On the role of risk-associated genetic loci in modulating clinical course in multiple sclerosis.

By

Gongbu Pan

B.A (Biological science; China)
M.S (Genetics; China)

Thesis submitted in fulfilment of the requirements for the degree of Master of Philosophy

University of Tasmania

June 2016

Declaration of originality

This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

Signature

	7th June 2016
Gongbu Pan	Date

Statement of authority to access

This thesis may be made available for loan. Copying and communication of any part of this thesis is prohibited for two years from the date this statement was signed; after that time limited copying and communication is permitted in accordance with the Copyright Act 1968.

Signature

	7th June 2016
Gongbu Pan	Date

Statement of Co-authorship

This thesis includes papers for which Gongbu Pan is not sole author. Gongbu Pan took the lead in this research, developing and implementing the analyses, writing manuscript included herein under the supervision of Bruce V Taylor (BVT), Ingrid van der Mei (IvM), Steve Simpson, Jr. (SSJ) and Jac Charlesworth (JC). In this process, however, he was assisted by supervisors to varying extent.

Signature	**Signature**
Professor Bruce Taylor	Professor Alison Venn
Supervisor	Director
Menzies Institute for Medical Research	Menzies Institute for Medical Research
University of Tasmania	University of Tasmania
Date 7th June 2016	Date

The paper reported in Chapter 1:

Gongbu Pan, Steve Simpson, Jr; Ingrid van der Mei, Jac C.Charlesworth,Yuan Zhou, Feitong Wu, Bruce Taylor. "The potential role of genetic factors in multiple sclerosis onset and clinical course: a narrative review".

The paper reported in Chapter 2:

Gongbu Pan, Steve Simpson, Jr; Ingrid van der Mei; Jac C. Charlesworth, Robyn Lucas, Anne-Louise Ponsonby, Yuan Zhou; Feitong Wu, AusLong/Ausimmune Investigator Group, Bruce Taylor. "Role of genetic susceptibility variants in predicting clinical course in multiple sclerosis: a cohort study" *Journal of Neurology, Neurosurgery & Psychiatry*.24 August 2016 0:1–8. doi:10.1136

Statement of Ethical Conduct

The research associated with this thesis abides by the international and Australian codes on human and animal experimentation, the guidelines by the Australian Government's Office of the Gene Technology Regulator and the rulings of the Safety, Ethics and Institutional Biosafety Committees of the University.

Signature

	7th June 2016
Gongbu Pan	Date

Acknowledgement

I would like to thank my supervisors, Bruce, Ingrid, Steve and Jac, for Bruce's vital encouragement and patient guidance, for Ingrid's invaluable advice and logical thinking, for Steve's generous assistance. I would especially like to thank my co-supervisor Steve. Thanks for his warm help on improving my English, quick answer and response, Stata coding guidance, useful comments and his kindness make me feel comfortable in Australia. They provided me the intellectual freedom to develop my own ideas. As a Chinese student, I need to thank all my supervisors' help for improving my English skills, writing and speaking, and their kindness makes me feel comfortable in Australia. I also want to thank my classmate, Feitong Wu and YuanZhou, for their assistance. Finally, I would like to thank my parents and Professor Changhai Ding, for their kind support and help.

Abstract

Multiple sclerosis (MS) is a chronic and immune-mediated central nervous system condition characterized by inflammation and demyelination. Large genome-wide association studies (GWAS) have identified 110 non-human leukocyte antigen (HLA) and several HLA MS risk variants. However, the relationship between MS risk-associated single nucleotide polymorphisms (SNPs) and MS clinical course has not been well studied. Despite good success in identifying a number of risk variants, several large GWAS studies have made comparatively little progress in finding associations with MS clinical course. Moreover, these risk SNPs only can explain 28% of MS risk. This brings one question: whether these risk SNPs influence MS clinical course?

This thesis first presents an overview of MS, including its clinical symptom & diagnosis, the epidemiology of MS, and particularly the role of genetic factors in MS. Following from this is the report of my work evaluating the role of MS risk-associated single nucleotide polymorphisms (SNPs) in modulating clinical course in early MS. Utilising a longitudinal cohort study of persons who have had a first demyelinating event suggestive of, but not yet diagnostic of MS and then followed for five years, we found:

- Seven non-HLA SNPs predicted relapse and/or CDMS, and seven other non-HLA SNPs predicted the annualised change in disability status (ΔEDSS, as measured by Expanded Disability Status Scale).
- Following from this, we generated two genetic risk scores (GRS) based on those identified risk SNPs associated with CDMS/relapse and disability progression, which each significantly predicted each outcome in a significant, dose-dependent manner:
    - Patients having >5 risk genotypes had a 6-fold greater risk of CDMS and relapse compared to those with <2;
    - Those carrying ≥6 risk genotypes progressed at 0.48 EDSS points per year faster compared to those with <2, and the CGRS model explained 32% of the variance in disability.

In summary, our findings of longitudinal data demonstrate that MS susceptibility genes predicted worse MS clinical course (that is, conversion to active disease, relapse risk and disability progression), suggesting that the genetic drivers of MS progression are polygenic. These findings may aid in targeting patients of high disease risk and potentially in early prevention and treatment, but replication is warranted.

Conference presentations arising from work in this thesis

Oral presentations

- 2015 Known multiple sclerosis genetic susceptibility variants are associated with clinical course: a cohort study. Progress in MS Research Conference 2015, in Melbourne.

Poster presentations:

- 2015 Graduate Research Conference Thursday, September 3rd, 2015 Hobart, University of Tasmania.
- 2015 5th Australia China Biomedical Research conference, Hobart Satellite Meeting, 2 November, 2015, Hobart, Australia.

List of abbreviations

MS	Multiple sclerosis
CNS	Central nervous system
FDE	First demyelinating event
CDMS	Clinically definite multiple sclerosis
EDSS	Expanded Disability Status Scale
MSSS	Multiple Sclerosis Severity Score
RRMS	Relapsing-remitting MS
PPMS	Primary-progressive MS
UVR	Ultraviolet radiation
25(OH)D	25-hydroxyvitamin D
$1,25(OH)_2D$	1,25-dihydroxyvitamin D
VDR	Vitamin D receptor
EBV	Epstein-Barr virus
MHC	Major histocompatibility complex
HLA	Human leukocyte antigen
GWAS	Genome-wide association studies
SNP	Single-nucleotide polymorphism
MAF	Minor allele frequency
LD	Linkage disequilibrium
IMSGC	International Multiple Sclerosis Genetics Consortium

Chapter 1. Introduction

1.1 Multiple Sclerosis

Multiple sclerosis (MS) is a chronic and immune-mediated central nervous system (CNS) condition characterized by inflammation and demyelination (1) with typical onset between 20 and 40 years old (2). While most patients present with the relapsing-remitting type, virtually all patients have some degree of disability progression over time, some catastrophically so. It is estimated that 2.3 million people are affected by MS globally, with a high prevalence (1/1,000) in Caucasian populations and with a high female/male sex ratio (1.5 to 2.5).

1.2 Clinical symptom and course of MS

Patients with MS can present with a range of symptoms, including limb weakness & muscle spasms, paraesthesia, vision problem (eye pain/dysfunction), balance issues, headache and ataxia, as well as bladder, bowel and sexual dysfunction; a majority of patients may also experience cognitive decline. Patients with MS may only have one or several of these symptoms at any one time and over the course of their disease.

1.2.1 MS Clinical course

MS clinical course comprises both onset (risk) and progression (relapse & disability progression). The components and stages of the disease in MS can be charactered as clinically isolated syndrome (CIS), first demyelinating event (FDE), clinically definite multiple sclerosis (CDMS), relapse and disability progression. The initial presentation in MS typically takes the form of a clinically isolated syndrome, usually a first attack of symptoms compatible with MS, such as optic neuritis, which while clinically significant is insufficient to merit a diagnosis of MS. Diagnosis of MS requires further evidence in support of actual disease, often in the form of a second episode substantiating dissemination of demyelination in time (serial and distinct episodes) and space (affecting different parts of the brain). This evidence

can also come from magnetic resonance imaging (MRI) lesions which can show dissemination in time and space, will have progression to CDMS (3-6).

The two key aspects of MS, at least clinically, are relapse and progression. A relapse (also known as an attack or exacerbation) refers to clinical symptoms or signs typically manifested as an acute inflammatory demyelinating event in the CNS(6). Clinically, a minimum duration of 24 hours for a relapse is typically required while in most cases this is much longer. Progression is characterized as a sustained loss of neurologic function or increased disability. It is from these two primary clinical elements of MS, relapse and disability progression, that define its three main clinical phenotypes: relapsing-remitting (RRMS), secondary progressive (SPMS), and primary progressive MS (PPMS) (Figure 1)(7). RRMS is the most common course of the disease. It is characterized by unpredictable but clearly defined relapses of new symptoms or increasing severity of existing symptoms, followed by periods of partial or total recovery (remissions). Of those RRMS patients, about two in three will develop to SPMS, a disease phase in which relapses cease but a greater degree of sustained disability progression occurs. A minority of MS patients, usually thought to be around 10%, present in the primary-progressive form, having no relapses during the course of their disease but instead suffering sustained disability progression from onset. The insidious nature of disability progression, particularly in patients who have no understanding of what is transpiring, makes the identification of the onset of this condition, and to a lesser extent its diagnosis, a more complicated affair in comparison to the relapsing-remitting type. In addition to these classic three types of MS is a somewhat less definite progressive-relapsing type (PRMS), which features the same insidious onset and progressive nature of PPMS, but with somewhat infrequent exacerbations of disability equated with relapses(8).

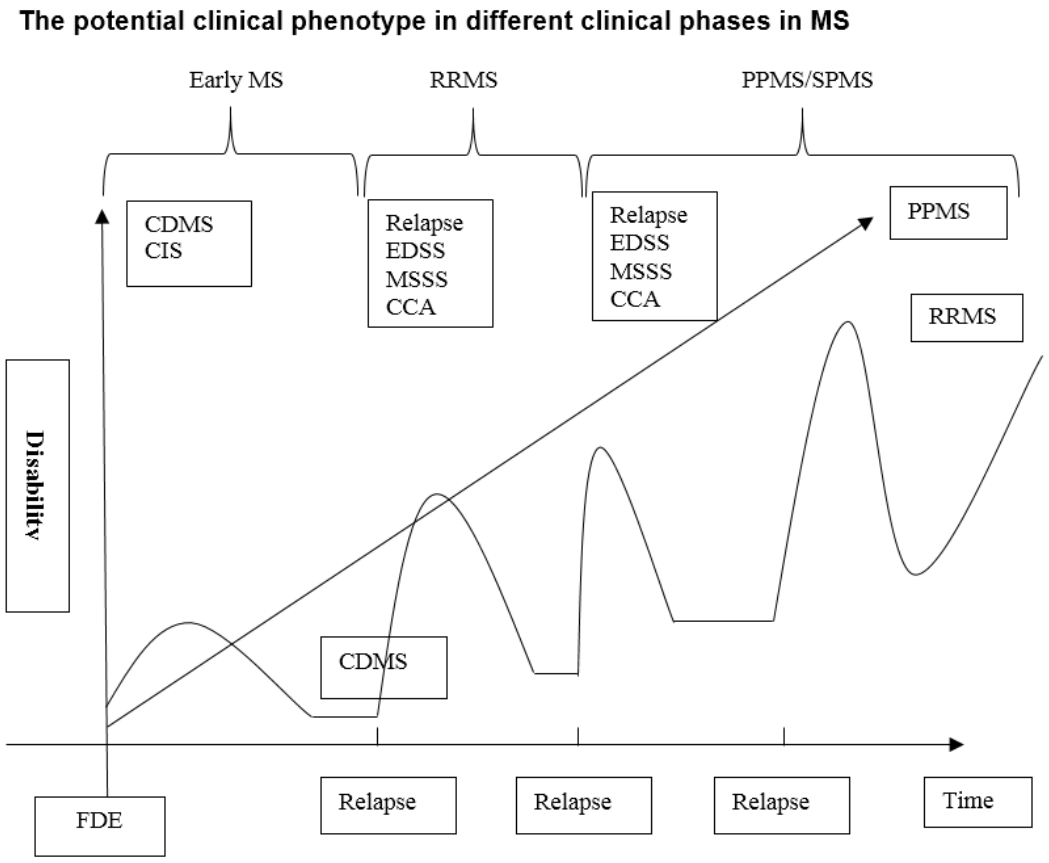


Figure 1 The potential clinical phenotype in different clinical phases in early MS. FDE: first demyelination episode. CDMS: clinically definite MS. CIS: clinically isolated syndrome. EDSS: Expanded Disability Status Scale. MSSS: Multiple Sclerosis Severity Score. RRMS: relapsing remitting MS. PPMS: primary progressive MS. SPMS: secondary progressive MS. CCA: Corpus callosum atrophy.

1.2.2 Measurement of disability progression

The most widely used method of quantifying disability in MS is the Expanded Disability Status Scale (EDSS) developed by a neurologist called John Kurtzke, based on the dysfunction of eight functional systems(9). The total EDSS scale ranges from 0 to 10 (normal to death), in a 0.5 unit increment (except the first interval which just goes from 0 to 1.0). Another measure of progression is the multiple sclerosis severity score (MSSS)(10) which adjusts the absolute level of disability by EDSS for the duration elapsed since symptom onset, such that persons with equivalent disability but the shorter duration will have a greater MSSS than someone with a longer duration. While relapses are acutely impactful on patient's quality of life, the sustained and irreversible nature of disability progression makes it highly correlated with quality of life.

1.3 Diagnosis and treatment of multiple sclerosis

The mean age of MS onset is presently thought to be around 30 years, and most people are diagnosed between 25-30 years old. Around 3-5% of diagnosed MS patients are under 18 years old (11, 12); a separate condition, paediatric-onset MS, presents before the age of 10. The main requirement for MS diagnosing is the demonstration of central nervous system disorder, based on a combination of clinical and MRI findings. Both physical and Magnetic resonance imaging (MRI) test are important for MS diagnosis. MRI always provide evidence to support the MS clinical diagnosis. The gold MS diagnostic criteria for MS, McDonald Criteria, also include specific criteria for MS MRI testing.

1.3.1 Gold standard for diagnosis of MS

Since it was first developed in 2001 by the International Panel on the Diagnosis of Multiple Sclerosis(4), the McDonald criteria has become the most popular one for MS diagnosis. This criteria requires clinical and laboratory assessments to indicate dissemination of CNS demyelinating lesions in space and in time, that is affecting different parts of the brain/spinal cord and being distinct episodes rather than one confluent episode. As newer evidence accumulates, this criteria has since been revised two times in 2005(5) and 2010(6), primarily to simplify the MRI diagnostic criteria while maintaining diagnostic sensitivity and specificity (Figure 1.2).

The 2010 McDonald Criteria for Diagnosis of MS

Clinical Presentation	Additional Data Needed for MS Diagnosis
≥2 attacks[a]; objective clinical evidence of ≥2 lesions or objective clinical evidence of 1 lesion with reasonable historical evidence of a prior attack[b]	None[c]
≥2 attacks[a]; objective clinical evidence of 1 lesion	Dissemination in space, demonstrated by: ≥1 T2 lesion in at least 2 of 4 MS-typical regions of the CNS (periventricular, juxtacortical, infratentorial, or spinal cord)[d]; or Await a further clinical attack[a] implicating a different CNS site
1 attack[a]; objective clinical evidence of ≥2 lesions	Dissemination in time, demonstrated by: Simultaneous presence of asymptomatic gadolinium-enhancing and nonenhancing lesions at any time; or A new T2 and/or gadolinium-enhancing lesion(s) on follow-up MRI, irrespective of its timing with reference to a baseline scan; or Await a second clinical attack[a]
1 attack[a]; objective clinical evidence of 1 lesion (clinically isolated syndrome)	Dissemination in space and time, demonstrated by: For DIS: ≥1 T2 lesion in at least 2 of 4 MS-typical regions of the CNS (periventricular, juxtacortical, infratentorial, or spinal cord)[d]; or Await a second clinical attack[a] implicating a different CNS site; and For DIT: Simultaneous presence of asymptomatic gadolinium-enhancing and nonenhancing lesions at any time; or A new T2 and/or gadolinium-enhancing lesion(s) on follow-up MRI, irrespective of its timing with reference to a baseline scan; or Await a second clinical attack[a]
Insidious neurological progression suggestive of MS (PPMS)	1 year of disease progression (retrospectively or prospectively determined) plus 2 of 3 of the following criteria[d]: 1. Evidence for DIS in the brain based on ≥1 T2 lesions in the MS-characteristic (periventricular, juxtacortical, or infratentorial) regions 2. Evidence for DIS in the spinal cord based on ≥2 T2 lesions in the cord 3. Positive CSF (isoelectric focusing evidence of oligoclonal bands and/or elevated IgG index)

If the Criteria are fulfilled and there is no better explanation for the clinical presentation, the diagnosis is "MS"; if suspicious, but the Criteria are not completely met, the diagnosis is "possible MS"; if another diagnosis arises during the evaluation that better explains the clinical presentation, then the diagnosis is "not MS."

[a]An attack (relapse; exacerbation) is defined as patient-reported or objectively observed events typical of an acute inflammatory demyelinating event in the CNS, current or historical, with duration of at least 24 hours, in the absence of fever or infection. It should be documented by contemporaneous neurological examination, but some historical events with symptoms and evolution characteristic for MS, but for which no objective neurological findings are documented, can provide reasonable evidence of a prior demyelinating event. Reports of paroxysmal symptoms (historical or current) should, however, consist of multiple episodes occurring over not less than 24 hours. Before a definite diagnosis of MS can be made, at least 1 attack must be corroborated by findings on neurological examination, visual evoked potential response in patients reporting prior visual disturbance, or MRI consistent with demyelination in the area of the CNS implicated in the historical report of neurological symptoms.

[b]Clinical diagnosis based on objective clinical findings for 2 attacks is most secure. Reasonable historical evidence for 1 past attack, in the absence of documented objective neurological findings, can include historical events with symptoms and evolution characteristics for a prior inflammatory demyelinating event; at least 1 attack, however, must be supported by objective findings.

[c]No additional tests are required. However, it is desirable that any diagnosis of MS be made with access to imaging based on these Criteria. If imaging or other tests (for instance, CSF) are undertaken and are negative, extreme caution needs to be taken before making a diagnosis of MS, and alternative diagnoses must be considered. There must be no better explanation for the clinical presentation, and objective evidence must be present to support a diagnosis of MS.

[d]Gadolinium-enhancing lesions are not required; symptomatic lesions are excluded from consideration in subjects with brainstem or spinal cord syndromes.

MS = multiple sclerosis; CNS = central nervous system; MRI = magnetic resonance imaging; DIS = dissemination in space; DIT = dissemination in time; PPMS = primary progressive multiple sclerosis; CSF = cerebrospinal fluid; IgG = immunoglobulin G.

Figure 1.2: McDonald criteria 2010 revisions (adapted from Polman et al.(6))

1.3.2 Magnetic resonance imaging support the diagnosis of MS

Magnetic resonance imaging (MRI), is a medical imaging test that is used to image the physiological processes of the brain. MRI has become the preferred imaging method to help diagnose MS and to monitor its clinical course, as highlighted by above-mentioned the McDonald diagnostic criteria for MS, which incorporate detailed standards for dissemination in space and time as demonstrated by MRI (Figure 1.2).

Cerebral or spinal plaques are the characteristic lesions for MS patients, showing as bright images on T2 weighted MRI. The plaque of MS focal lesions can be neat-edged and round, oval or irregular in appearance and arranged at right angles to the cerebral hemisphere and corpus callosum on MRI. When viewed on MRI sagittal images, a typical element is the Dawson finger (Figure 1.3). Dawson finger is a demyelinating plaque through the corpus callosum. Patients can have one or most of these characteristics. The size of a plaque varies, usually from a few millimeters to 1cm. Plaques can be found in the temporal lobe, internal capsule, cortex, and cerebellum region. Over the disease course, the cumulative impacts of neuroinflammation can manifest as a degeneration and atrophy, particularly evident in patients with severe MS. The frequency of spinal cord lesions are less common in European-descent populations, but more common in Asian populations (13). They are more commonly found in the cervical and thoracic spine on MRI, and usually located in the outer periphery of the spinal cord white matter, mostly seen as a small plaque. The size of these plaques is at least at 3mm but not more than two vertebral segments in length(14).(Figure 1.4)

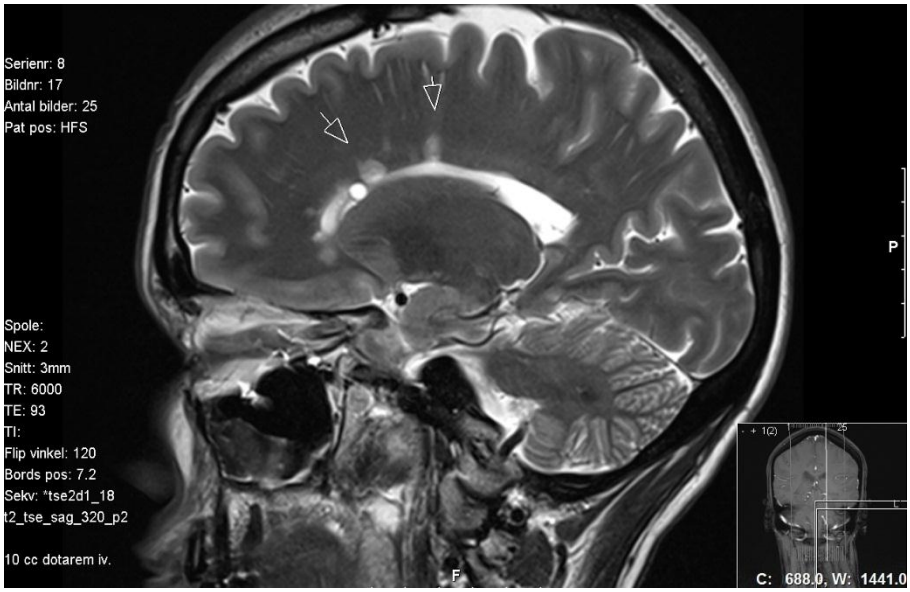


Figure 1.3: The Dawson Fingers (From Radiology MRI http://radiologymri.blogspot.com.au/2012/01/proton-density-in-multiple-sclerosis.html)

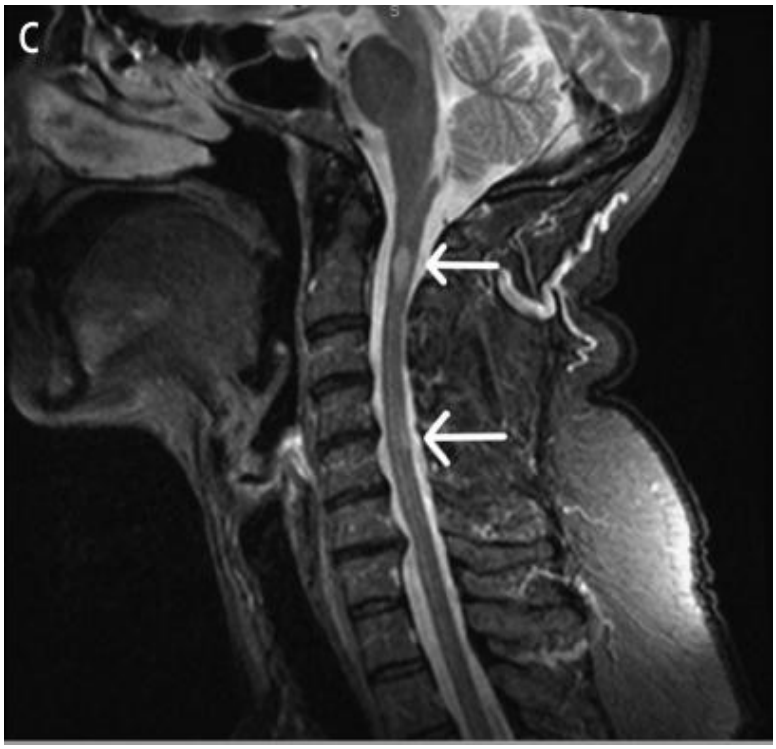


Figure 1.4: T2-weighted sagittal view of spinal cord, which shows a hyperintense lesion in cervical spinal cord(15).

1.3.3 Treatment of MS

As the cause of MS remains unclear, as yet there is no cure. Treatments only can relieve symptoms and delay the progression of MS (16). Methylprednisolone can be taken as infusions or pills and is used to reduce neuroinflammation. It is useful in controlling the severity of MS attacks. Interferon-beta is one kind of useful treatment for MS. Interferon-beta is derived from the human cytokine that modulates immune responsiveness(17). Interferon-beta medications are injected, either intramuscular or subcutaneous and act to reduce the severity and frequency of relapses in relapsing-remitting MS patients. Other medications such as glatiramer acetate which can prevent immune system's attack on myelin are also widely used. Monoclonal antibodies (MAB) are one of the newest treatment for MS. This treatment is widely used to modulate the immune system by working on targets involved in MS development. This treatment seems to be more useful in both now and future. However, there is strong interest in developing drugs for progressive MS with several studies now showing some effect on preventing disability accumulation in primary progressive MS (Ocrelizumab) and in secondary progressive MS (glutathione and simvastatin), although reversal of disability has not been shown.

1.4 The epidemiology of MS

In the past several decades of research, epidemiology has made essential contributions to our understanding of MS, including its geographic distribution, its distribution in society, as well as identifying a number of significant risk or protective environmental factors, behaviors and genetic susceptibility variants.(18-21).

1.4.1 The geo-epidemiology of MS

Previous studies have shown a latitudinal gradient of MS prevalence, with a high prevalence of MS observed in areas of high latitude. Geographically, MS divided into several major

frequency areas(22), particularly in the most part of Europe, Canada, the northern United States and southeastern Australia (ranging from 58-116/100,000); in many parts of these areas, the prevalence is more than 100/100,000; the Orkney Islands have the highest global prevalence of 300/100,000(23). The southern United States, South Africa, northern Australia, most parts of Russia and some parts of Latin America have a prevalence ranging from 10 to 60/100,000, representing medium frequency areas. Southeast Asia and Africa have the lowest prevalence, generally between 0 and 10. These data, derived from countless local and regional epidemiology studies, are summarised in the "Atlas of MS", an online and text-based resource produced by the MS International Federation. The 2013 Atlas of MS reported a total number of persons with MS globally of nearly 2.3 million, and showed a significant sex ratio in MS prevalence, with the female prevalence more than three times that of males. Western and European-descent countries have the highest prevalence estimates of MS, according to the latest data shown in Figure 1.5. Although MS is found worldwide, it is less common in Asian populations and quite rare in Aboriginal populations such as the Inuit, Māori, Eskimo and Australian Aborigines. The mean age of MS onset is nearly 30 years, though studies indicate the mean age of onset is a few years earlier in females than males. According to the latest Survey of Disability, Ageing & Carers (SDAC) in 2009, there were 23,700 Australians affected by MS. The national prevalence was 95.2/100.000, though the frequency is much greater at higher latitudes like Tasmania than northern latitudes like Queensland (24).

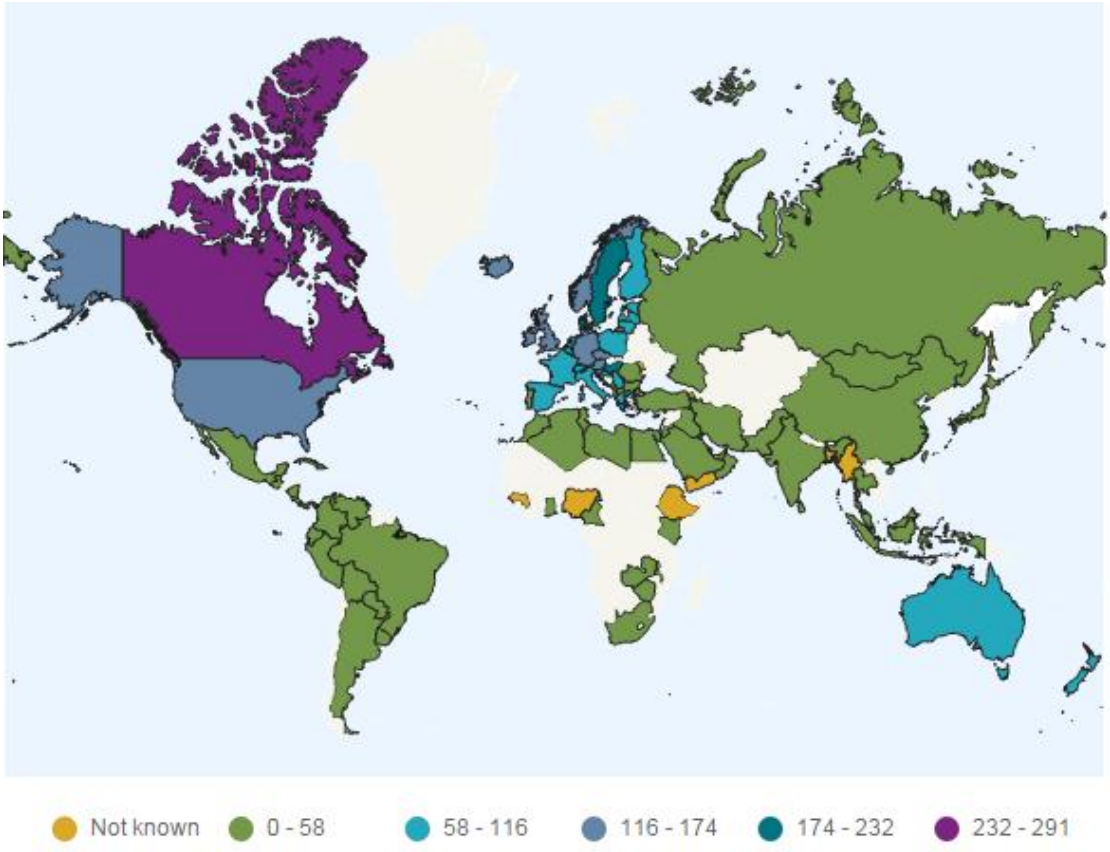


Figure 1.5: The worldwide prevalence of MS (adapted from Atlas of MS 2013).

1.5 The environmental factors associated with MS

With the development of MS pathogenic mechanisms research, considerable evidence indicates that environmental factors play an important role in the aetiology of MS. Although all the environmental factors affecting MS development have not yet been established, accumulating evidence has strongly implicated several environmental factors, including Epstein-Barr virus (EBV) and human herpesvirus 6 (HHV-6) infection, vitamin D deficiency and low ultraviolet (UV) exposure, and tobacco smoking.

1.5.1 Virus infection (EBV & HHV-6)

EBV, is one of the most common human viruses, it is a double-stranded DNA γ-herpesvirus. This virus manifests as a latent infection within B-cells after they are infected (25). It has been demonstrated that EBV infection is associated with MS risk, with almost a hundred

percent detection rate in MS patients (26). Higher rates of EBV seropositivity have been identified in MS patients comparing with controls(27). However, infectious mononucleosis history is also a risk factor for MS, which supports the EBV association. In a study of US military personnel, the risk of MS among those who had anti-EBNA titres > 320 units was 36-times higher than those who had titres < 20 (28). Of note, infection with EBV during adolescence showed a further 2-3 fold increased the risk of developing MS(29). However, one study found that the reactivation parameters of EBV were not associated with MS clinical course, and this is supported by a subsequent independent 5-yr prospective study, which did not show any associations between EBV levels and MS activity and progression(30). These findings may suggest that while some MS environmental risk factors are also associated with MS progression, others may only influence the onset.

Human herpesvirus-6 (HHV-6; two types, i.e., HHV-6A and HHV-6B), first discovered by Salahuddin and colleagues in 1986(31). There is increasing evidence demonstrating that this virus has a potential role in the pathogenesis of MS. It is found that one in four of MS patients' saliva and plasma had HHV-6B DNA (32). Although this virus can also be found in plasma from healthy people, low MS disease activity patients had a lower frequency of HHV-6B positive samples than those having high disease activity (32). Moreover, the reactivation of HHV-6A has been associated with two genes (IRF5 and MHC2TA rs4774C), which have been previously related to increased MS susceptibility (33). Importantly, serum DNA levels of HHV-6 have been reduced after treatment with interferon beta, while patients had poor outcomes (e.g., more relapses and a lesser reduction in relapse rate) when their HHV-6 DNA levels were not reduced by therapy (34). Nevertheless, more longitudinal studies are needed to confirm the association between HHV-6 and MS risk and clinical course, and RCTs are needed to explore useful treatment for MS targeted at clearing this virus.

1.5.2 Vitamin D & UV exposure

Vitamin D deficiency has been considered as a risk factor for many complex diseases such as MS. Recent studies have shown that low vitamin D levels play an important role on MS risk. A large number of epidemiological data indicated that there is a high prevalence of MS in high-latitude and low ultraviolet radiation (UVR) areas, in contrast to the low prevalence of MS in low-latitude and high UVR areas. This may be because vitamin D_3 is produced photochemically from precursors in the skin by UVR exposure(35). In the liver, vitamin D_3 converted to the circulating form, 25-hydroxyvitamin D, and then converted to 1,25-dihydroxyvitamin D in kidney cells, as well as in other cell types, particularly neuroglial and immune system cells. Vitamin D play an essential role in modulating the immune response.

Now, it is well established that Vitamin D in humans mainly comes from sun exposure. Vitamin D play an essential role in the synthesis of myelin. Vitamin D as an environmental factor, may not only influence whether a person will get MS but also may impact the clinical course of MS. The involvement of vitamin D in MS onset and progression was related to the essential role that vitamin D pathway plays in the autoimmune system. The immune system are regulated by combining of the active form vitamin D (hydroxylates 25-hydroxyvitamin D) with the specific vitamin D receptor (*VDR*); and the function of *VDR* is regulated by its genetic structure. There are many restriction sites such as Bsm I and Apa I in the *VDR* gene. One study suggests that the polymorphisms of the *VDR* gene may be associated with MS risk(36). However, this finding was not confirmed by a meta-analysis(37).

Simpson and colleagues found that each 10nmol/l increase in vitamin D resulted in a nearly 12% reduction of MS relapse risk by using survival analysis(38). This is the first longitudinal cohort study to investigate the relationship between vitamin D and relapse of RRMS patients. The recent subsequent study shows evidence that Vitamin D plays an important role in myelin repair. This study finds vitamin D activated RXR gamma receptor protein and this protein was important in MS patients' myelin repairing (39). This finding functionally explained the relationship between the vitamin D level and progression of MS and indicated that vitamin D

could be a target for the future myelin repair drug and a possible treatment for people with MS. In short, compared to other MS treatment, vitamin D based treatment are cheaper with fewer side effects. It is easier for people to increase vitamin D level by sun exposure and daily intake.

1.5.3 Tobacco smoking

Tobacco smoking has been considered as a risk factor of MS for many years. In addition, recent studies have demonstrated that tobacco smoking influence both risk of developing MS and MS progression (relapse and disability)(40, 41). Studies have indicated that those tobacco smokers have a greater risk of MS compared to those non-smokers; and MS risk is also associated with the duration and intensity of smoking (42, 43). One recent study has shown patients who smoked at least 10 cigarettes per day had a higher risk than never-smokers of progressing to secondary progressive MS(40). While the mechanism by which smoking is related to MS onset and progression remains uncertain, it has been suggested that this may be partly explained by inflammation or immune dysfunction (44).

1.6 Genetic research in MS

Although the aetiology of MS is not yet fully understood, genetic factors have been shown to play a crucial role in MS risk. However, little is known about the role of genetic factors in MS clinical course, including clinically definite MS (CDMS), relapse and disability progression (2, 45). In the last two decades, a total of 110 non-HLA region genetic loci and several HLA genetic loci associated with MS susceptibility have been identified by genome-wide association studies (GWAS)(46). However, only 28% of the perceived heritability of MS can be explained by those genetic loci and large GWAS have not found any associations between these MS susceptibility genes and MS clinical course. Therefore, what explains the missing heritability and whether genetic factors have potential roles in modulating MS clinical course needs to be answered.

In addition to the previously mentioned potential role of differential UVR exposure in mediating the latitudinal variation in MS, some studies have suggested a role for genetic factors due to differential migration and genetic risk. The incidence of MS in Western countries is higher than that in Asian countries. Kira found that MS incidence rate is very low in low-latitude regions, such as South America and Africa, which is related to the low frequency of *HLA-DR2* gene(47). Some ethnic groups, such as Eskimos, Aboriginal populations in North America, Australia & New Zealand, and Romani, have lower frequencies of MS, suggesting a role for genetic factors in MS susceptibility.

The hypothesis that genetic factors may be the main determinants of MS familial aggregation was first implied by the observation of clustering of MS in families based on the results of the previous family and twin studies. In line with previous work(48), the Canada twin study demonstrated the concordance rate in monozygotic twin (25.3%) was four times more than that seen in dizygotic twin(5.4%)(49). It has also been suggested that the children's recurrence risk is influenced by both parents' genetic background (50). These data, long recognised and continuing to the present, provided the initial evidence in support of a genetic component to MS, and that this effect is polygenic. However, even while there is a recognised genetic element to MS, the fact that the concordance rate in monozygotic twin is only 30% suggests that environmental exposures also influence individual's risk to MS.

1.6.1 Linkage and Candidate analysis in MS genetic research

The main findings of this period are the HLA II area associated with MS onset which was confirmed by linkage and candidate gene approaches. Based on these findings, linkage and candidate gene approach were used to estimate the effect of substantial MS risk genes. Linkage analysis is a powerful tool based on the findings that genes located physically close together on a chromosome during the meiosis phase are inherited together as a block.(51) Linkage analysis became the primary mode of statistical genetic mapping of Mendelian and

complex diseases with familial aggregation in early genetic research. Linkage study based on family data has been useful to examine the effect of specific genes associated with Mendelian (i.e., monogenic) diseases, but it has been less effective in identifying susceptibility genes of polygenic complex disease such as MS. In 2007, the International Multiple Sclerosis Genetics Consortium (IMSGC) established the largest non-parametric genome-wide linkage screen in 931 MS family trios. The important result for this study was identifying the significant linkage within the HLA locus, with a peak in the regional logarithm of the odds (LOD) score of 11.7, while none of the loci outside the HLA locus reached statistical significance. Then, lots of studies investigated genetic factors involved in MS onset, however no validated new susceptibility genes were identified. The *HLA-DRB1*15:01* allele had the greatest effect on risk of that region area. Generally, given the complexity of MS, and the potential for gene-environment interaction, linkage and candidate gene studies have not been successful at finding susceptibility loci outside the MHC region in MS. Partly this is because these studies are limited by small sample size, as well as the small number of candidate variants available for previous analyses.

1.6.2 Genome-wide association studies and success for MS genetic research

Researchers found that common variants in many genes outside the HLA region will each have small effects on the risk of complex diseases such as MS (OR up to 1.30 for non-HLA SNPs (52)). According to these findings, the focus of genetic studies has shifted from linkage analysis to association analysis of the common variants in complex disease because association analysis studies are more successful than linkage methods in searching disease risk variants. Thus, automated chip based GWAS testing using millions of genetic variation marker, the single nucleotide polymorphism (SNP) was developed , and has become the favoured genetic mapping method to investigate the genetic structure of complex disease (53). Many GWAS have often used a multistage design. In Stages One and Two, well-established case and control samples are genotyped by GWAS chips. A representative Cartesian plot

needs to be presented to show the DNA were successfully genotyped. In the quality control and data cleaning stage, microarray quality, genotype completion rates, Hardy-Weinberg equilibrium analysis, sex inconsistency tests, parent offspring genotype incompatibilities, principal components analysis and removing bad variants and poor performing subjects need to be undertaken. Finally, statistical analysis is performed by testing individual SNPs for association by using appropriate regression models for quantitative traits. The statistical significance (P-value) must be adjusted for the number of tests to reduce the risk of false positives. A threshold of $5.0x10^{-8}$ is typically defined as genome-wide significant.

Although GWAS have discovered many SNPs predicting a number of conditions (54), some major challenges have been raised.

1) **Ethnicity mix**, referred to as population stratification, is a particular kind of confounding. People from different ancestries have a different allele frequency in some SNPs (55). GWAS can be confounded by population stratification which can lead to an increasing number of false positive results. The way to correct population stratification bias is to measure this confounder and adjust for it in regression models, or to stratify the cases and controls removing outliers at the time of principal components analysis.

2) **Multiple comparisons adjustment.** Statistical significance is usually defined as $p<0.05$ in classic epidemiological studies. However, for GWAS, the significance threshold must be adjusted to account for the substantial number of tests for SNPs. Bonferroni correction is widely used for multiple comparison adjustment. The Bonferroni method sets GWAS statistical significance at $p=0.05/N$. However, this method is sometimes considered to be too strict a criterion because some tests are not independent. In GWAS, many SNPs are in linkage disequilibrium (LD), which means they are correlated. To deal with this, the high number of test needs to be replaced. There are two common approaches to deal with this problem: 1) to calculate the true

number of independent tests(56); 2) to use the false discovery rate (FDR) method to adjust the FDR for an independent test, which has been shown to be effective at improving statistical power (57). These methods suggest a genome significant threshold between 10^{-6} and 10^{-8}, but the classic GWAS threshold of 10^{-8} remains widely used.

The first MS GWAS paper was published in 2007 by IMSGC(58) (Table 1). This paper, testing 334,923 SNPs in 931 cases, identified polymorphisms in *HLA-DRA* locus ($p=8.94x10^{-81}$), *IL2RA* ($P=2.96x10^{-8}$) and *IL7RA* ($p=2.94x10^{-7}$) as risk factors for MS. The *IL7RA* gene finding was subsequently replicated in other MS studies(59, 60).This paper opened a new era in MS genetic research. Since this publication, 16 genome-wide association studies have been published (Table 1).The first large MS GWAS study in Australia and New Zealand(61) was published by the ANZGene Collaboration, which established a genome-wide association study of 1,618 cases and 3,413 controls. They identified six new susceptibility variants on chromosomes 12 and 20. This study also confirmed several previously identified MS risk variants.

Table 2: List of major GWAS in MS genetic research

	Year	Population	Number of cases	Number of controls
IMSGC	2007	USA,UK	931	1,862
WTCCC1	2007	UK	975	1,466
Comabella	2008	Spain	242	242
Aulchenko	2008	Netherlands	45	195
Baranzini	2009	Various	978	883
De Jager	2009	USA	860	1,720
ANZgene	2009	AU, New Zealand	1,618	3,413
Sanna	2010	Sardinia	882	872
Nischwitz	2010	Germany	590	825

Jakkula	2010	Finland	68	136
IMSGC,WTCCC2	2011	Various	9,772	17,376
Patsopoulos	2011	Various	1453	2176
Matesanz	2012	Spain	296	801
Martinelli	2012	Italy	197	234
IMSGC	2013	European	14,498	24,091
IMSGC	2015	European	17,465	30,385

1.6.3 HLA loci associated with MS

The HLA region of the genome encodes the major histocompatibility complex (MHC). MHC region is a key element of adaptive immunity and spans a region of about 4,000Kb located on the short arm of chromosome 6 at band position 6p21.3. It is the most complex genetic polymorphism system of the human genome. The MHC region is divided into three major regions: HLA class I, encoding the MHC I proteins; class II, encoding the MHC II proteins; and class III, which are partly of unknown function but appear to play some role in the general immune function and regulation(62). The HLA class I and class II region are the most polymorphic region in human DNA sequences, leading to a great variety of genotypes, this a valuable component of adaptive immune recognition of the diverse number of epitopes to be recognised. The class I region contains *HLA-A*, *HLA-B* and *HLA-C* genes, while the class II region contains *HLA-DP*, *HLA-DQ* and *HLA-DR* genes. Variants in this regions have been associated with most inflammatory and autoimmune disease such as multiple sclerosis.

Bertrams and colleagues identified that *HLA-A3* and *HLA-B7* were associated with MS onset(63). Subsequent study has confirmed this finding(64). Altogether, HLA variants have explained 10.5% of the heritability of MS(65). Of the HLA loci, most studies have shown the HLA-DRB1*1501 locus to be the most strongly associated with MS risk (OR=3.1, $P_{combined}<10^{-320}$)(66-68), connoting a three-fold greater risk of MS(69). Several other HLA alleles and haplotypes have also been identified as risk factors of MS onset, but these appear

to vary in their association with risk by ancestry and ethnicity. Stankovich and colleagues investigated HLA associations in 1,230 MS cases in Australia, finding both *HLA-DRB1*1501* and *HLA-DRB1*03* to be associated with MS risk (70). In Turkey and the Canary Islands, *HLA-DRB1*04* was identified as a risk factors of MS(71), while in Sardinia, *HLA-DRB1*0301* and *HLA-DRB1*0401* were identified as genetic risk factors(71). MS patients in Canada and Sweden were associated with *HLA-DRB1*17*(72). *HLA-DRB1*1501* and *HLA-DQB1*0602* have an effect on in persons of African descent (67). Finally, even though Asian MS patients typically present as optico-spinal MS (OS-MS), in contrast to that seen in European-descent populations, MS patients in Japan also were associated with *HLA-DRB1*1501*(73); and the incidence of MS patients with *HLA-DRB1*0405* was significantly increased(74). Field and colleagues identified a relationship between MS and *HLA DRB1*15:01, *0301, *0401, *1303, HLA-A*0201* and *HLA-DPB1*0301* in 1,618 MS case and *3,413* controls of European descent (75). A SNP in an intron of *HLA-DPB1*, rs3135021, was also associated with MS in African Americans (68). In the same year, a Polish group identified that a polymorphism in the *HLA-G* gene was a risk factor for MS patients (76). In 2012, Alcina and colleagues reported that the A allele of SNP rs3135388 in *DRB1*1501* was associated with *DQB1, DRB1* and *DRB5* genes' high expression in a Caucasian compared to G allele (77). These data demonstrate a significant role for the *DRB1*1501* locus in MS onset. Not only are these loci associated with MS onset, but also there is some evidence showing that deleterious HLA genotypes like *DRB1*1501* were positively associated with a marker of disease, oligoclonal bands (OB), while protective variants like *DRB1*0405* were negatively associated with OB (78). These findings are useful in that they provide internal consistency and validation to HLA being associated with MS. Another study reported that the late onset MS was significantly associated with *HLA-DRB1*0801* gene (79). In 2007, Silva reported *DRB1*15* allele may be associated with a better MS clinical course (80). In 2009, Rama and other studies have found that *HLA-DR15*1501* has a significant effect on the age of onset(AO) of MS among Caucasians and African-Americans (81). Figure 1.6 shows the HLA region's genetic map and the risk of multiple sclerosis.

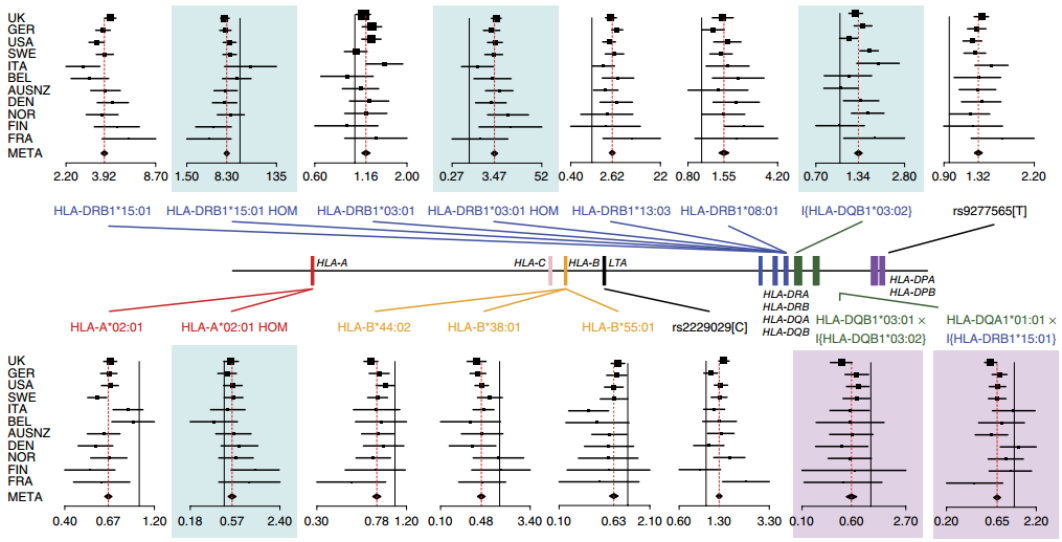


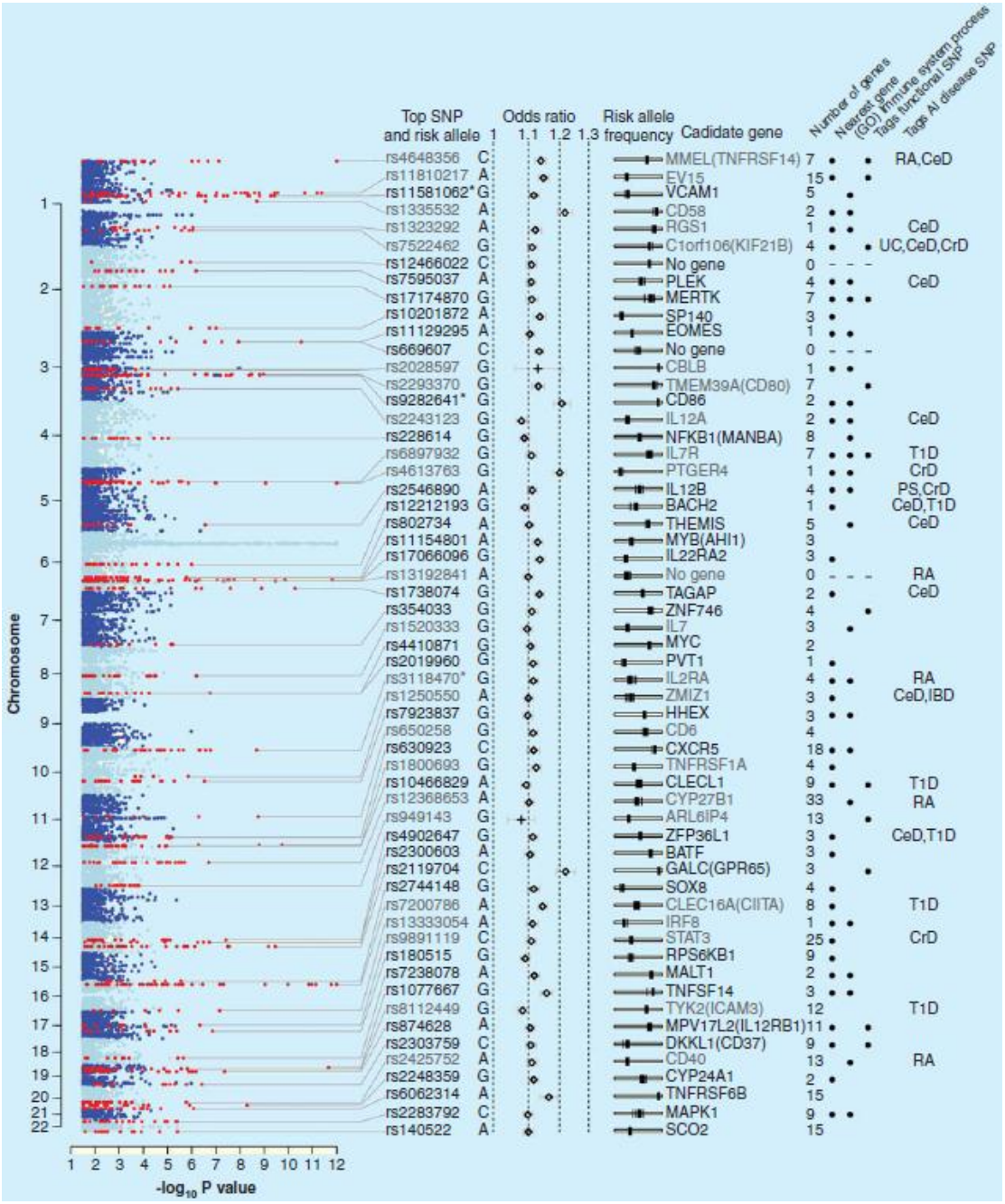
Figure Overview of genetic effects in the HLA region influencing risk for multiple sclerosis. The relative locations of the classical HLA loci are shown (central bar) along with forest plots for each of the major effects identified. Each forest plot shows the estimated odds ratio from the cohort-specific logistic model and the 95% confidence interval, with point size proportional to sample size and the cohorts ordered by size The bottom point for each locus and the corresponding red dashed line show the result from fixed-effect meta-analysis (META). Plot background indicates the nature of the effect: none, main effect; blue, departure from additivity; pink, interaction. The lead allele for each signal is indicated. The solid black line represents OR = 1 for additive effects and the expected homozygote effect for departures from additivity. Curly brackets represent an indicator function that combines heterozygotes and homozygotes into a single category. HOM, homozygote.

Figure 1.6 Genetic map of the HLA region and the risk of multiple sclerosis. (reproduced from Moutsianas and colleagues(82))

1.6.4 110 non-HLA common variants associated with MS

During the last few decades, utilising candidate-gene, linkage studies, and GWAS approaches, at least 110 non-HLA genetic loci have been identified as associated with MS onset(52, 69, 83-86) (Appendix 1). In contrast, there has been comparatively less work into the genetic drivers of MS clinical course. The large GWAS have shown no significant loci that differentiate progressive-onset from bout-onset MS, even in cohorts enriched for progressive cases(61). Similarly, no association has been found with disability(84, 85, 87, 88).This likely reflects the comparative difficulty in evaluating clinical course in genetic studies since MS clinical course (conversion to active disease, relapse, or disability progression) is not easily studied by GWAS, as GWAS are cross-sectional or case-control in design. While MS clinical course is best assessed longitudinally, and ideally in real time and from disease onset, so as to reduce potential impacts of reverse causality or heterogeneity by treatment or other disease aspects. Most of these 110 variants' biological pathway still unclear, as yet, there are no benefits for MS treatment, based on GWAS result. In addition,

few relationships between these associated variants and MS progression were identified. One Western Australian study identified that *HLA-DRB1*1501* was correlated with MS severity as measured by MSSS, and *HLA-DRB1*1201* was correlated with less severe of MS (89). Our own studies also found some SNPs influenced MS clinical course (see Chapter 2).



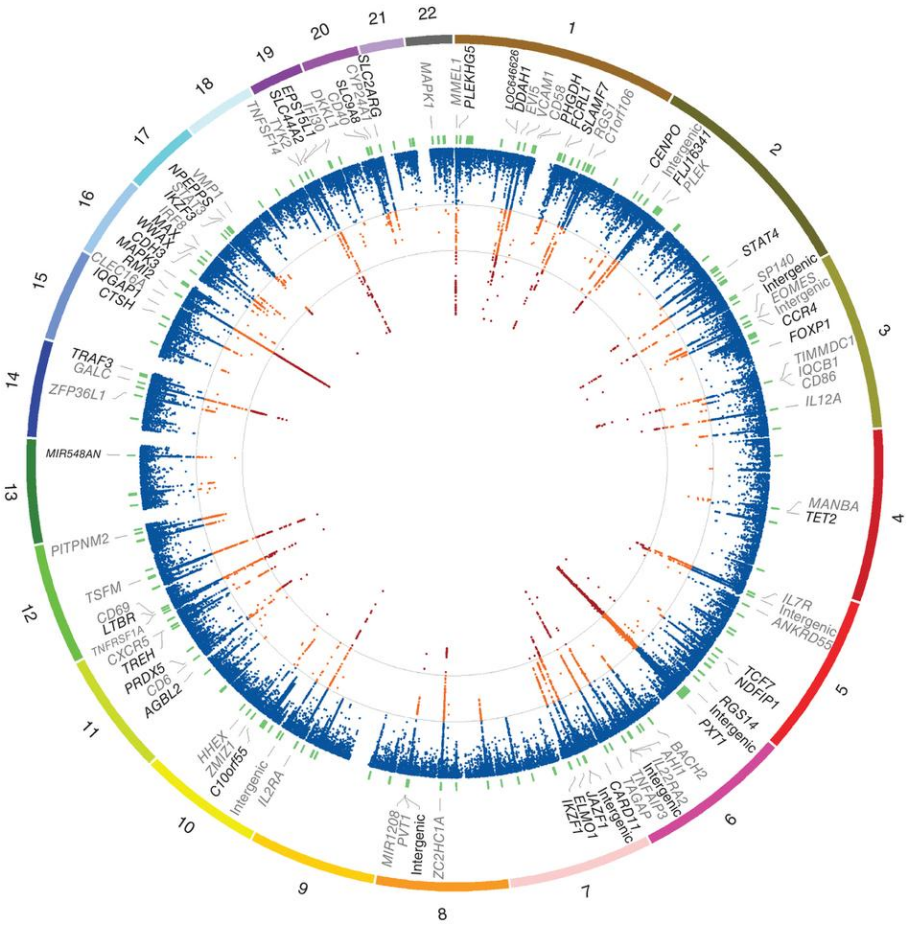


Figure 1.7: The genomic map of 110 non HLA loci in Multiple Sclerosis (reported by IMSGC)

1.7 The effect of genetic on MS clinical course

1.7.1 Gene-environment interactions

Studies have shown that environmental factors and genetic factors are the two principal risk factors for multiple sclerosis. The most well-known environmental risk factors for MS are vitamin D deficiency, smoking and EBV infection as discussed above(90). Studies have indicated that gene-environment interactions can explain a part of the MS genetic missing heritability.

Lin and colleagues used Cox proportional hazards regression models to estimate the relationship between known MS common risk variants and MS relapse; and whether these

SNPs influence the 25-hydroxyvitamin D (25(OH)D) –relapse association among 141 RRMS patients in the Southern Tasmanian Multiple Sclerosis Longitudinal (MSL) Study(45). They identified five variants associated with MS relapse with significant cumulative genotype risk effects, but also found three SNPs which modified the relationship between the hazard of relapse and serum 25(OH)D levels. Unfortunately, no risk variants were identified for MS disability progression. This study indicated that gene-environment interactions may play an important role in MS clinical course. It will provide support for the role of serum 25(OH)D in MS onset and progression. Similarly, another study using the same cohort showed protein kinase C (PKC) family genes-25(OH)D interactions modulate MS clinical course(91). In this study, they identified two SNPs tagging *PRKCZ* and *PRKCH* that interacted with 25(OH)D levels to influence relapse. However, the SNPs themselves were not independently associated with hazard of relapse. They also found two SNPs within the *CYP2R1* and *PRKCB* that were associated with 25(OH)D levels and relapse. *CYP2R1* has been previously identified to be associated with vitamin D levels in GWAS. The *PRKCB* gene has been previously associated with risk of rheumatoid arthritis and it is a member of the PKC gene family which is regulated by1,25 dihydroxyvitamin D (1,25(OH)D) in chondrocytes and mediated by *VDR* in downstream signalling pathways. These results suggest that gene-vitamin D interaction could be associated with MS clinical course.

Similarly, the rate of progression of established MS is highly variable, e.g. monozygotic twins can have onsets of MS at significantly different ages, have completely different clinical presentations, and can progress at very different rates. This variability may be under genetic control or at least influenced by it. Currently, no genes have been found that influence different MS phenotype (PPMS vs. RRMS) or severity. The *HLA-DRB1*1501* locus decreases the age of onset marginally (less than 2 years) (92). Therefore, it is likely that genes that interact with environmental factors (vitamin D, EBV) may significantly influence the rate of MS progression and determine the different MS phenotype. It is thought that secondary progressive MS and PPMS represent the same pathological process and that MS relapses may

represent a separate and additive inflammatory process more directly influenced by environmental determinants.

Many research questions arise from the observation that a proportion of genetically susceptible individuals remain healthy while others develop MS. One potential answer is genetic heterogeneity, and the other is gene-environmental interactions. Genetic epidemiology implies that genetic background has an important complementary role in MS onset. As genetic factors remain the same in individuals, the environment determines the threshold of MS onset(93).

1.7.2 Epigenetic contributions to MS

It has been suggested that epigenetic mechanism may contributes to the pathophysiology of MS(94, 95). Epigenetics may alter MS risk gene's expression. These changes may potentially mediate the response to environmental influences and result in functionally significant changes in gene expression.

Epigenetics is defined as non- heritable changes in gene expression but not modifying the underlying DNA sequence(96). The epigenetic mechanisms consist of histone modification, DNA methylation and miRNA-associated post-transcriptional gene silencing(97). The major changes of epigenetic processes in cells are DNA methylation and histone deacetylation. That MS is usually transmitted to children more by female than male supports an epigenetic contribution(98).

Epigenetic changes could explain much of the heterogeneity in MS clinical course. For example, gene methylation of MS patients' peripheral blood mononuclear cells (PBMCs) found that DNA methylation may be used as markers of MS disease activity (99). A study on methylation changes in MS patients tested 56 candidate genes and found that 15 genes in the cell-free plasma DNA show methylation changes in MS patients compared to the healthy controls group, further, 5 of the 15 differentially methylated genes may also distinguish

patients with remission and exacerbation(99). One immunology study found that T(H)-17 cell-associated MicroRNA-326 was correlated with MS severity. This means MicroRNA-326 plays a critical role in T(H)-17 differentiation and could be associated with MS onset. A subsequent study found that MicroRNA-155 promotes the progression of autoimmune inflammation by enhancing T cell development (100). Another study showed that miRNA-572 was significantly upregulated and downregulated during MS clinical course such as disease relapse and remission phases. This result indicated that miRNA-572 may be a biomarker for remyelination and modulates MS relapse and remission(101). These results suggest that microRNAs might be potential targets in the treatment of MS. However, significant further investigation of the role MicroRNA in MS treatment is needed.

Though the epigenetics contribution to MS is just beginning to be understood, several mechanisms of epigenetic change in patients with MS have been identified. No doubt there will be more discoveries in the future. It has opened a new era in MS research. DNA demethylation of epigenetic, histone deacetylation and regulation mechanism of miRNA offer enticing prospect of new therapies for MS. There is a vast potential to use epigenetics for both MS preventive treatments and personalised therapies. It also brings a great hope to develop complex diseases' molecular targeted genome research (102, 103).

References

1. Lassmann H, Bruck W, Lucchinetti CF. The immunopathology of multiple sclerosis: an overview. Brain Pathol. 2007;17(2):210-8.
2. Sawcer S, Franklin RJ, Ban M. Multiple sclerosis genetics. Lancet Neurol. 2014;13(7):700-9.
3. Berger T, Rubner P, Schautzer F, Egg R, Ulmer H, Mayringer I, et al. Antimyelin antibodies as a predictor of clinically definite multiple sclerosis after a first demyelinating event. N Engl J Med. 2003;349(2):139-45.
4. McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Ann Neurol. 2001;50(1):121-7.
5. Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". Ann Neurol. 2005;58(6):840-6.
6. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol. 2011;69(2):292-302.
7. Lublin FD, Reingold SC, Cohen JA, Cutter GR, Sorensen PS, Thompson AJ, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. Neurology. 2014;83(3):278-86.
8. Lublin FD, Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. Neurology. 1996;46(4):907-11.
9. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology. 1983;33(11):1444-52.
10. Roxburgh RH, Seaman SR, Masterman T, Hensiek AE, Sawcer SJ, Vukusic S, et al. Multiple Sclerosis Severity Score: using disability and disease duration to rate disease severity. Neurology. 2005;64(7):1144-51.
11. Boiko A, Vorobeychik G, Paty D, Devonshire V, Sadovnick D, University of British Columbia MSCN. Early onset multiple sclerosis: a longitudinal study. Neurology. 2002;59(7):1006-10.
12. Bigi S, Banwell B. Pediatric multiple sclerosis. J Child Neurol. 2012;27(11):1378-83.
13. Bot JC, Barkhof F, Polman CH, Lycklama a Nijeholt GJ, de Groot V, Bergers E, et al. Spinal cord abnormalities in recently diagnosed MS patients: added value of spinal MRI examination. Neurology. 2004;62(2):226-33.
14. Bag AK, Patel BN, Osman S, Roberson GH. Clinico-radiologic profile of spinal cord multiple sclerosis in adults. Neuroradiol J. 2011;24(4):511-8.
15. Klawiter EC. Current and new directions in MRI in multiple sclerosis. Continuum (Minneap Minn). 2013;19(4 Multiple Sclerosis):1058-73.
16. Nylander A, Hafler DA. Multiple sclerosis. J Clin Invest. 2012;122(4):1180-8.
17. Kasper LH, Reder AT. Immunomodulatory activity of interferon-beta. Ann Clin Transl Neurol. 2014;1(8):622-31.
18. Simpson S, Jr., Taylor BV, van der Mei I. The role of epidemiology in MS research: Past successes, current challenges and future potential. Mult Scler. 2015;21(8):969-77.

19. Koch-Henriksen N, Sorensen PS. The changing demographic pattern of multiple sclerosis epidemiology. Lancet Neurol. 2010;9(5):520-32.
20. Alonso A, Hernan MA. Temporal trends in the incidence of multiple sclerosis: a systematic review. Neurology. 2008;71(2):129-35.
21. Orton SM, Herrera BM, Yee IM, Valdar W, Ramagopalan SV, Sadovnick AD, et al. Sex ratio of multiple sclerosis in Canada: a longitudinal study. Lancet Neurol. 2006;5(11):932-6.
22. Kurtzke JF. Epidemiologic contributions to multiple sclerosis: an overview. Neurology. 1980;30(7 Pt 2):61-79.
23. Handel AE, Jarvis L, McLaughlin R, Fries A, Ebers GC, Ramagopalan SV. The epidemiology of multiple sclerosis in Scotland: inferences from hospital admissions. PLoS One. 2011;6(1):e14606.
24. Simpson S, Jr., Blizzard L, Otahal P, Van der Mei I, Taylor B. Latitude is significantly associated with the prevalence of multiple sclerosis: a meta-analysis. J Neurol Neurosurg Psychiatry. 2011;82(10):1132-41.
25. Pender MP, Burrows SR. Epstein-Barr virus and multiple sclerosis: potential opportunities for immunotherapy. Clin Transl Immunology. 2014;3(10):e27.
26. Ascherio A, Munger KL, Lunemann JD. The initiation and prevention of multiple sclerosis. Nat Rev Neurol. 2012;8(11):602-12.
27. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part I: the role of infection. Ann Neurol. 2007;61(4):288-99.
28. Munger KL, Levin LI, O'Reilly EJ, Falk KI, Ascherio A. Anti-Epstein-Barr virus antibodies as serological markers of multiple sclerosis: a prospective study among United States military personnel. Mult Scler. 2011;17(10):1185-93.
29. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part I: the role of infection. Ann Neurol. 2007;61(4):288-99.
30. Munger KL, Fitzgerald KC, Freedman MS, Hartung HP, Miller DH, Montalban X, et al. No association of multiple sclerosis activity and progression with EBV or tobacco use in BENEFIT. Neurology. 2015;85(19):1694-701.
31. Salahuddin SZ, Ablashi DV, Markham PD, Josephs SF, Sturzenegger S, Kaplan M, et al. Isolation of a new virus, HBLV, in patients with lymphoproliferative disorders. Science. 1986;234(4776):596-601.
32. Hollsberg P, Kusk M, Bech E, Hansen HJ, Jakobsen J, Haahr S. Presence of Epstein-Barr virus and human herpesvirus 6B DNA in multiple sclerosis patients: associations with disease activity. Acta Neurol Scand. 2005;112(6):395-402.
33. Dominguez-Mozo MI, Garcia-Montojo M, De Las Heras V, Garcia-Martinez A, Arias-Leal AM, Casanova I, et al. MHC2TA mRNA levels and human herpesvirus 6 in multiple sclerosis patients treated with interferon beta along two-year follow-up. BMC Neurol. 2012;12:107.
34. Garcia-Montojo M, De Las Heras V, Dominguez-Mozo M, Bartolome M, Garcia-Martinez MA, Arroyo R, et al. Human herpesvirus 6 and effectiveness of interferon beta1b in multiple sclerosis patients. Eur J Neurol. 2011;18(8):1027-35.
35. Holick MF. Vitamin D: A millenium perspective. J Cell Biochem. 2003;88(2):296-307.
36. Niino M, Fukazawa T, Yabe I, Kikuchi S, Sasaki H, Tashiro K. Vitamin D receptor gene polymorphism in multiple sclerosis and the association with HLA class II alleles. J Neurol Sci. 2000;177(1):65-71.
37. Huang J, Xie ZF. Polymorphisms in the vitamin D receptor gene and multiple sclerosis risk: a meta-analysis of case-control studies. J Neurol Sci. 2012;313(1-2):79-85.

38. Simpson S, Jr., Taylor B, Blizzard L, Ponsonby AL, Pittas F, Tremlett H, et al. Higher 25-hydroxyvitamin D is associated with lower relapse risk in multiple sclerosis. Ann Neurol. 2010;68(2):193-203.
39. de la Fuente AG, Errea O, van Wijngaarden P, Gonzalez GA, Kerninon C, Jarjour AA, et al. Vitamin D receptor-retinoid X receptor heterodimer signaling regulates oligodendrocyte progenitor cell differentiation. J Cell Biol. 2015;211(5):975-85.
40. Roudbari SA, Ansar MM, Yousefzad A. Smoking as a risk factor for development of Secondary Progressive Multiple Sclerosis: A study in IRAN, Guilan. J Neurol Sci. 2013;330(1-2):52-5.
41. Pittas F, Ponsonby AL, van der Mei IA, Taylor BV, Blizzard L, Groom P, et al. Smoking is associated with progressive disease course and increased progression in clinical disability in a prospective cohort of people with multiple sclerosis. J Neurol. 2009;256(4):577-85.
42. Hernan MA, Olek MJ, Ascherio A. Cigarette smoking and incidence of multiple sclerosis. Am J Epidemiol. 2001;154(1):69-74.
43. Hedstrom AK, Baarnhielm M, Olsson T, Alfredsson L. Tobacco smoking, but not Swedish snuff use, increases the risk of multiple sclerosis. Neurology. 2009;73(9):696-701.
44. Costenbader KH, Karlson EW. Cigarette smoking and autoimmune disease: what can we learn from epidemiology? Lupus. 2006;15(11):737-45.
45. Lin R, Taylor BV, Simpson S, Jr., Charlesworth J, Ponsonby AL, Pittas F, et al. Association between multiple sclerosis risk-associated SNPs and relapse and disability--a prospective cohort study. Mult Scler. 2014;20(3):313-21.
46. Psychiatric GCBDWG. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat Genet. 2011;43(10):977-83.
47. Kira J. Multiple sclerosis in the Japanese population. Lancet Neurol. 2003;2(2):117-27.
48. Sadovnick AD, Armstrong H, Rice GP, Bulman D, Hashimoto L, Paty DW, et al. A population-based study of multiple sclerosis in twins: update. Ann Neurol. 1993;33(3):281-5.
49. Willer CJ, Dyment DA, Risch NJ, Sadovnick AD, Ebers GC, Canadian Collaborative Study G. Twin concordance and sibling recurrence rates in multiple sclerosis. Proceedings of the National Academy of Sciences of the United States of America. 2003;100(22):12877-82.
50. Robertson NP, O'Riordan JI, Chataway J, Kingsley DP, Miller DH, Clayton D, et al. Offspring recurrence rates and clinical characteristics of conjugal multiple sclerosis. Lancet. 1997;349(9065):1587-90.
51. Pulst SM. Genetic linkage analysis. Arch Neurol. 1999;56(6):667-72.
52. International Multiple Sclerosis Genetics C, Beecham AH, Patsopoulos NA, Xifara DK, Davis MF, Kemppinen A, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat Genet. 2013;45(11):1353-60.
53. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nature reviews Genetics. 2005;6(2):95-108.
54. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic acids research. 2014;42(Database issue):D1001-6.

55. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. Nature reviews Genetics. 2010;11(7):459-63.
56. Li MX, Yeung JM, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. Hum Genet. 2012;131(5):747-56.
57. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300.
58. International Multiple Sclerosis Genetics C, Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, et al. Risk alleles for multiple sclerosis identified by a genomewide study. N Engl J Med. 2007;357(9):851-62.
59. Lundmark F, Duvefelt K, Iacobaeus E, Kockum I, Wallstrom E, Khademi M, et al. Variation in interleukin 7 receptor alpha chain (IL7R) influences risk of multiple sclerosis. Nat Genet. 2007;39(9):1108-13.
60. Gregory SG, Schmidt S, Seth P, Oksenberg JR, Hart J, Prokop A, et al. Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis. Nat Genet. 2007;39(9):1083-91.
61. Australia, New Zealand Multiple Sclerosis Genetics C. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. Nat Genet. 2009;41(7):824-8.
62. Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, et al. Gene map of the extended human MHC. Nat Rev Genet. 2004;5(12):889-99.
63. Bertrams HJ, Kuwert EK. Association of histocompatibility haplotype HLA-A3-B7 with multiple sclerosis. J Immunol. 1976;117(5 Pt.2):1906-12.
64. Wadia NH, Trikannad VS, Krishnaswamy PR. HLA antigens in multiple sclerosis amongst Indians. J Neurol Neurosurg Psychiatry. 1981;44(9):849-51.
65. Hollenbach JA, Oksenberg JR. The immunogenetics of multiple sclerosis: A comprehensive review. J Autoimmun. 2015;64:13-25.
66. Olerup O, Hillert J. HLA class II-associated genetic susceptibility in multiple sclerosis: a critical evaluation. Tissue Antigens. 1991;38(1):1-15.
67. Oksenberg JR, Barcellos LF, Cree BA, Baranzini SE, Bugawan TL, Khan O, et al. Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. Am J Hum Genet. 2004;74(1):160-7.
68. McElroy JP, Cree BA, Caillier SJ, Gregersen PK, Herbert J, Khan OA, et al. Refining the association of MHC with multiple sclerosis in African Americans. Hum Mol Genet. 2010;19(15):3080-8.
69. International Multiple Sclerosis Genetics C, Wellcome Trust Case Control C, Sawcer S, Hellenthal G, Pirinen M, Spencer CC, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011;476(7359):214-9.
70. Stankovich J, Butzkueven H, Marriott M, Chapman C, Tubridy N, Tait BD, et al. HLA-DRB1 associations with disease susceptibility and clinical course in Australians with multiple sclerosis. Tissue Antigens. 2009;74(1):17-21.
71. Dyment DA, Ebers GC, Sadovnick AD. Genetics of multiple sclerosis. Lancet Neurol. 2004;3(2):104-10.
72. Modin H, Olsson W, Hillert J, Masterman T. Modes of action of HLA-DR susceptibility specificities in multiple sclerosis. Am J Hum Genet. 2004;74(6):1321-2.
73. Fukazawa T, Yamasaki K, Ito H, Kikuchi S, Minohara M, Horiuchi I, et al. Both the HLA-CPB1 and -DRB1 alleles correlate with risk for multiple sclerosis in Japanese:

clinical phenotypes and gender as important factors. Tissue Antigens. 2000;55(3):199-205.
74. Yoshimura S, Isobe N, Yonekawa T, Matsushita T, Masaki K, Sato S, et al. Genetic and infectious profiles of Japanese multiple sclerosis patients. PLoS One. 2012;7(11):e48592.
75. Field J, Browning SR, Johnson LJ, Danoy P, Varney MD, Tait BD, et al. A polymorphism in the HLA-DPB1 gene is associated with susceptibility to multiple sclerosis. PLoS One. 2010;5(10):e13454.
76. Wisniewski A, Bilinska M, Klimczak A, Wagner M, Majorczyk E, Nowak I, et al. Association of the HLA-G gene polymorphism with multiple sclerosis in a Polish population. Int J Immunogenet. 2010;37(4):307-11.
77. Alcina A, Abad-Grau Mdel M, Fedetz M, Izquierdo G, Lucas M, Fernandez O, et al. Multiple sclerosis risk variant HLA-DRB1*1501 associates with high expression of DRB1 gene in different human populations. PLoS One. 2012;7(1):e29819.
78. Wu JS, Qiu W, Castley A, James I, Joseph J, Christiansen FT, et al. Presence of CSF oligoclonal bands (OCB) is associated with the HLA-DRB1 genotype in a West Australian multiple sclerosis cohort. J Neurol Sci. 2010;288(1-2):63-7.
79. Qiu W, Wu JS, Castley A, James I, Joseph J, Christiansen FT, et al. Clinical profile and HLA-DRB1 genotype of late onset multiple sclerosis in Western Australia. J Clin Neurosci. 2010;17(8):1009-13.
80. Silva AM, Pereira C, Bettencourt A, Carvalho C, Couto AR, Leite MI, et al. The role of HLA-DRB1 alleles on susceptibility and outcome of a Portuguese Multiple Sclerosis population. J Neurol Sci. 2007;258(1-2):69-74.
81. Ramagopalan SV, Byrnes JK, Dyment DA, Guimond C, Handunnetthi L, Disanto G, et al. Parent-of-origin of HLA-DRB1*1501 and age of onset of multiple sclerosis. J Hum Genet. 2009;54(9):547-9.
82. Moutsianas L, Jostins L, Beecham AH, Dilthey AT, Xifara DK, Ban M, et al. Class II HLA interactions modulate genetic risk for multiple sclerosis. Nat Genet. 2015;47(10):1107-13.
83. International Multiple Sclerosis Genetics C, International IBDGC, International IBDGCI. Class II HLA interactions modulate genetic risk for multiple sclerosis. Nat Genet. 2015.
84. Baranzini SE, Wang J, Gibson RA, Galwey N, Naegelin Y, Barkhof F, et al. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. Hum Mol Genet. 2009;18(4):767-78.
85. Martinelli-Boneschi F, Esposito F, Brambilla P, Lindstrom E, Lavorgna G, Stankovich J, et al. A genome-wide association study in progressive multiple sclerosis. Mult Scler. 2012;18(10):1384-94.
86. Patsopoulos NA, Barcellos LF, Hintzen RQ, Schaefer C, van Duijn CM, Noble JA, et al. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. PLoS Genet. 2013;9(11):e1003926.
87. Jensen CJ, Stankovich J, Van der Walt A, Bahlo M, Taylor BV, van der Mei IA, et al. Multiple sclerosis susceptibility-associated SNPs do not influence disease severity measures in a cohort of Australian MS patients. PLoS One. 2010;5(4):e10003.
88. van der Walt A, Stankovich J, Bahlo M, Taylor BV, van der Mei IA, Foote SJ, et al. Apolipoprotein genotype does not influence MS severity, cognition, or brain atrophy. Neurology. 2009;73(13):1018-25.
89. Wu JS, James I, Qiu W, Castley A, Christiansen FT, Carroll WM, et al. HLA-DRB1 allele heterogeneity influences multiple sclerosis severity as well as risk in Western Australia. J Neuroimmunol. 2010;219(1-2):109-13.

90. Lincoln JA, Cook SD. An overview of gene-epigenetic-environmental contributions to MS causation. J Neurol Sci. 2009;286(1-2):54-7.
91. Lin R, Taylor BV, Simpson S, Jr., Charlesworth J, Ponsonby AL, Pittas F, et al. Novel modulating effects of PKC family genes on the relationship between serum vitamin D and relapse in multiple sclerosis. J Neurol Neurosurg Psychiatry. 2014;85(4):399-404.
92. Simon KC vdMI, Munger KL, et al. Combined effects of smoking, anti-EBNA antibodies, and HLADRB1*1501 on multiple sclerosis risk. Neurology. 2010;74:1365–1371.

93. Ebers GC. Environmental factors and multiple sclerosis. Lancet Neurol. 2008;7(3):268-77.
94. Graves MC, Benton M, Lea RA, Boyle M, Tajouri L, Macartney-Coxson D, et al. Methylation differences at the HLA-DRB1 locus in CD4+ T-Cells are associated with multiple sclerosis. Mult Scler. 2014;20(8):1033-41.
95. Qureshi IA, Mehler MF. Epigenetic mechanisms underlying the pathogenesis of neurogenetic diseases. Neurotherapeutics. 2014;11(4):708-20.
96. Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. Nature. 2004;429(6990):457-63.
97. Koch MW, Metz LM, Kovalchuk O. Epigenetic changes in patients with multiple sclerosis. Nat Rev Neurol. 2013;9(1):35-43.
98. Ebers GC, Sadovnick AD, Dyment DA, Yee IM, Willer CJ, Risch N. Parent-of-origin effect in multiple sclerosis: observations in half-siblings. Lancet. 2004;363(9423):1773-4.
99. Liggett T, Melnikov A, Tilwalli S, Yi Q, Chen H, Replogle C, et al. Methylation patterns of cell-free plasma DNA in relapsing-remitting multiple sclerosis. J Neurol Sci. 2010;290(1-2):16-21.
100. O'Connell RM, Kahn D, Gibson WS, Round JL, Scholz RL, Chaudhuri AA, et al. MicroRNA-155 promotes autoimmune inflammation by enhancing inflammatory T cell development. Immunity. 2010;33(4):607-19.
101. Mancuso R, Hernis A, Agostini S, Rovaris M, Caputo D, Clerici M. MicroRNA-572 expression in multiple sclerosis patients with different patterns of clinical progression. J Transl Med. 2015;13:148.
102. Zhou Y, Simpson S, Jr., Holloway AF, Charlesworth J, van der Mei I, Taylor BV. The potential role of epigenetic modifications in the heritability of multiple sclerosis. Mult Scler. 2014;20(2):135-40.
103. Kucukali CI, Kurtuncu M, Coban A, Cebi M, Tuzun E. Epigenetics of multiple sclerosis: an updated review. Neuromolecular Med. 2015;17(2):83-96.

Chapter 2. Known multiple sclerosis genetic susceptibility variants are associated with clinical course: a cohort study

2.1 Abstract

Background

The genetic drivers of multiple sclerosis (MS) clinical course are essentially unknown with limited data arising from severity and clinical phenotype analyses in genome-wide association studies (GWAS).

Methods

Prospective cohort study of 127 first demyelinating events with genotype data, where 116 MS risk-associated single nucleotide polymorphisms (SNPs) were assessed as predictors of conversion to MS, relapse, and annualised disability progression (EDSS) up to 5-year review (ΔEDSS). Survival analysis was used to test for predictors of MS & relapse, and linear regression for disability progression. The top seven SNPs predicting MS/relapse and disability progression were evaluated as a cumulative genetic risk score (CGRS).

Results

We identified two non-HLA (rs12599600 & rs1021156) and one HLA (rs9266773) SNP predicting both CDMS and relapse risk. Additionally, three non-HLA SNPs predicted only conversion to MS; one HLA and two non-HLA SNPs predicted only relapse; and seven non-HLA SNPs predicted ΔEDSS. The CGRS significantly predicted CDMS and relapse in a significant, dose-dependent manner: those having $\geq$5 risk genotypes had a 6-fold greater risk of CDMS and relapse compared to those with $\leq$2. The CGRS for ΔEDSS was also significant:

those carrying ≥6 risk genotypes progressed at 0.48 EDSS points per year faster compared to those with ≤2, and the CGRS model explained 32% of the variance in disability in this study cohort.

Conclusions

These data strongly suggest that MS genetic risk variants significantly influence MS clinical course and that this effect is polygenic.

Keywords

Multiple sclerosis, epidemiology, genetics, relapse, disability, EDSS, cohort, longitudinal study.

2.2 Introduction

During the last few decades, utilising candidate-gene, linkage studies and genome-wide association study (GWAS) approaches, at least 6 human leukocyte antigen (HLA) loci and 110 non-HLA genetic loci have been identified as associated with multiple sclerosis (MS) onset (1-6). In contrast, there has been comparatively less work into the genetic drivers of MS clinical course. The large GWAS have shown no significant loci that differentiate progressive-onset from bout-onset MS, even in cohorts enriched for progressive cases(7). Similarly no association has been found with disability(4, 5, 8, 9).This likely reflects the comparative difficulty in evaluating clinical course in genetic studies, since MS clinical course (conversion to active disease, relapse, or disability progression) is not easily studied by

GWAS, as GWAS are cross-sectional or case-control in design. While MS clinical course is best assessed longitudinally, and ideally in real time and from disease onset, so as to reduce potential impacts of reverse causality or heterogeneity by treatment or other disease aspects.

A more expeditious approach to assess genetic determinants of clinical course is to utilise established GWAS determined MS onset associated variants, and evaluate these as predictors of MS clinical course in a prospective longitudinal cohort study. As we can hypothesize that those genetic variants associated with MS onset are potentially also involved in clinical course. This brings to bear the strengths of this study design, while mitigating the power limitations attendant to using a genome-wide approach.

Using this approach we have shown in a well-described longitudinal cohort of established MS cases, we have shown some evidence that known MS risk SNPs influence relapse and disability.(10) Here we extend this approach to analyse data from a prospective cohort of cases recruited around their first clinical episode suggestive of CNS inflammatory demyelination referred to as a first demyelinating event (FDE), and followed for five years with repeated neurological review. All measures of MS clinical course have been collected prospectively, including conversion to MS, relapse and measures of disability.

2.3 Methods

Study design

The Ausimmune case-control study(11), was designed to elucidate environmental and genetic risk factors for the onset and early progression of MS. The Ausimmune Study recruited a study sample of 282 case participants with a first clinical diagnosis of CNS demyelination indicating a high risk of developing MS. Case participants in the Ausimmune Study have been followed up in the AusLong cohort Study; (the analyses presented here) includeing follow-up to five years from study recruitment (84.6% retention).

The AusLong Study cohort included in these analyses is slightly different to the original Ausimmune Study case participant sample, as a result of clinical information provided up to the 5-year review. Three Ausimmune Study case participants were identified as not having had a MS-associated FDE (one neuromyelitis optica, one Susac's Syndrome, and one pineal germinoma). Additionally, 3 cases originally regarded as bout-onset were reclassified as being progressive-onset after follow-up.

The Ausimmune Study and AusLong Study were approved by nine regional Human Research Ethics Committees. All participants gave written informed consent.

2.4 Measurement of clinical outcomes

Several clinical outcomes were evaluated, including time to conversion to clinically definite MS (CDMS), number of relapses, and annualised disability progression from FDE to five year review (average 5.8 years from onset). Conversion to MS was defined primarily as the occurrence of two or more clinical demyelinating episodes, thus satisfying the diagnostic requirements of dissemination in space and time, or a single episode plus paraclinical

evidence, as per the 2005 McDonald criteria(12) (a minority of cases were diagnosed following MRI (either at the 2/3-year or 5-year reviews) based on this latter criterion (n=20)). Conversion to MS was reported at annual review and cross-checked with neurological records. A relapse was defined according to the 2001 McDonald Criteria(13) as the acute or subacute appearance or reappearance of a neurological abnormality (lasting at least 24 hours) in the absence of other potential explanatory factors. Relapses were reported at annual review and only relapses which were diagnosed and verified by a neurologist were included in this analysis. Disability was assessed by the Kurtzke Expanded Disability Status Scale (EDSS)(14) assessed at the 5-year review by the study neurologists.

2.5 Genotyping and SNP selection

DNA from AusLong participants was genotyped using the Illumina Human Exome BeadChip (Illumina Human Exome-12 v1.2 array) which includes ~244,000 exome SNPs with an additional ~87,000 MS relevant variants added as a customised component. Quality control(15) was conducted based on previous protocols. In general, individuals were excluded based on the following criteria: a call rate of <99%, gender error or duplicate discordance. Variants were excluded on the basis of a call rate of <99% or a deviation from Hardy-Weinberg equilibrium with $P<1.0x10^{-6}$. Principal components analysis was carried out to identify population outliers(16). Data on the previously published 110 MS-associated non-HLA region SNPs(1, 2, 7) and 6 HLA SNPs(2, 3, 17) were extracted for analysis.

2.6 Data analysis

Predictors of time to conversion to MS and to relapse were evaluated by Cox proportional hazards regression models, the latter for repeated events using the gap-time model by Prentice(18). All covariates satisfied the proportional hazards assumption.

Annualised change in EDSS (ΔEDSS) was calculated by taking the 5-year review EDSS and dividing by the duration between the day before the date of the FDE (EDSS assumed to be 0) and the 5-year review; this proportion was rendered into an annualised value. Predictors of ΔEDSS were evaluated using linear regression, adjusted for whether persons were having a relapse at the time of their 5-year EDSS assessment. Because the annualised change in disability was highly skewed, a log-transformation was applied to satisfy linear regression assumptions of minimal heteroskedasticity. All means and coefficients, however, were back-transformed and presented on the original scale of ΔEDSS. Adjustment for Bonferroni multiple comparison was applied for 116 SNPs (110 non-HLA and 6 HLA), this defined as the as-measured p-value multiplied by the number of tests (n=116)(19).

We created a cumulative genetic risk score (CGRS) which included the significant SNPs from the MS/relapse analysis and the ΔEDSS analysis separately. The common disease-common variant hypothesis predicts that common disease-causing alleles, or variants, will be found in all human populations which manifest a given disease. So we wanted to see how these

common variants predicted MS relapse and disability progression. We created two variables that provided values for the number of risk genotypes affecting outcomes, to represent two cumulative genetic risk scores(20-22). For example, those subjects with three, four or five genotypes that associated with higher probability of conversion to MS were each compared with the reference group – those carrying fewer than two associated SNPs. Where only the homozygous level of the risk genotype was significantly associated with outcomes, this was defined as the risk genotype, but where both the heterozygote and homozygote carriers of the risk genotypes were significantly associated with outcomes, these were defined as the risk genotypes.

To assess potential type 1 error and provide additional evidence to support that our findings did reflect altered risk of the outcome, we undertook a simulation involving the 14 SNPs found to significantly predict CDMS/relapse and disability progression (7 for CDMS/relapse, 7 for disability progression). For this analysis, a permutation simulation was done wherein AusLong participants' genotype data for these SNPs was randomly reallocated in equivalent proportions of genotype to that in the original sample. For example, the proportions of genotype rs842639 were such that 125 persons had the reference genotype and the remainder the non-reference genotype (Table 3). The simulated genotypes were generated, analysed and the magnitudes of the estimates resultant therefrom retained. These simulations were run 50,000 times and the proportion of magnitudes resulting that were as or more extreme than that found in the as-measured analyses denoted the significance for each SNP.

While the total study sample was 279 participants, the analyses in this paper are restricted to the 127 cases with a classic FDE and genotyping data for CDMS/relapse and 125 cases for disability progression. Additionally, 23 of the cases had PPMS.

All statistical analyses above were conducted in Stata/SE 12.1 (StataCorp LP, College Station, Texas, USA).

2.7 Results

Characteristics of participants

Of the 279 participants in the AusLong Study, genotype data were available for 207 participants; 127 of these had a classic FDE and were evaluated in our analyses. Of these, 98 (77.2%) were female and the mean age at study entry was 37.8 (SD: 9.5) years. 68 (53.5%) had converted to CDMS by 5-year review and had 151 relapses, while the median 5-year EDSS was 1 (interquartile range: 0-2). Of the 207 cases, 125 participants had an EDSS recorded at 5 years.

Non-HLA SNP predictors of clinical outcomes

We identified five non-HLA SNPs which predicted conversion to MS, while four non-HLA SNPs predicted relapse (Table 1). Two SNPs (rs1021156 near *PKIA* and *ZC2HC1A*, rs12599600 near *PRM1* and *RMI2*) were associated with both CDMS and relapse. None of the SNPs which predicted conversion to MS and/or relapse showed any association with ΔEDSS.

While none of these associations persisted in significance on adjustment for multiple comparisons (116 tests), the consistent effect direction between conversion to MS and relapse, even for those SNPs that did not significantly associated with the other outcome, increases our confidence that the associations are genuine.

Table 1 Seven top non HLA-SNPs and their associations with the hazard of conversion to MS and relapse*

SNP	Chr	Gene$	MS		Relapse	
			Number of CDMS (%)	HR (95% CI)	Number of Relapses (%)	HR (95% CI)
rs12599600#	16	*RMI2,*				
CC^		*PRM1*	43 (63.24)	1.00 [Reference]	107 (70.86)	1.00 [Reference]
CA+AA			25 (36.76)	**0.41 (0.24, 0.70)**	44 (29.14)	**0.54 (0.34, 0.87)**
Trend:				**p=0.001**		**p=0.011**
rs1021156	8	*ZC2HC1A,*				
CC		*PKIA*	33 (48.53)	1.00 [Reference]	69 (45.70)	1.00 [Reference]
CT^			28 (41.18)	1.44 (0.86, 2.42)	59 (39.07)	1.22 (0.77, 1.95)
TT^			7 (10.29)	**3.56 (1.96, 6.48)**	23 (15.23)	**2.41 (1.46, 3.97)**
Trend:				**p=0.003**		**p=0.015**
rs694739	11	*PRDX5,*				
AA^		*CCDC88B*	31 (45.59)	1.00 [Reference]	69 (45.70)	1.00 [Reference]
AG			32 (47.06)	0.69 (0.41, 1.18)	66 (43.71)	0.78 (0.51, 1.19)
GG			5 (7.35)	**0.34 (0.14, 0.83)**	16 (10.60)	0.71 (0.29, 1.76)
Trend:				**p=0.012**		p=0.31
rs802734	6	*PTPRK ,*				
AA		*THEMIS*	31 (45.59)	1.00 [Reference]	72 (47.68)	1.00 [Reference]
AG			25 (36.76)	0.89 (0.52, 1.51)	49 (32.45)	0.78 (0.47, 1.28)
GG^			12 (17.65)	**3.97 (1.83, 8.62)**	30 (19.87)	1.61 (0.96, 2.69)
Trend:				**p=0.034**		p=0.42
rs1359062#	1	*RGS1,*				
AA		*RGS21*	41 (60.29)	1.00 [Reference]	88 (58.28)	1.00 [Reference]
AG^+GG^			27 (39.71)	**1.71 (1.03, 2.84)**	63 (41.72)	1.41 (0.91, 2.20)
Trend:				**p=0.039**		p=0.13
rs35929052#	16	*IRF8,*				
CC^		*LOC146513*	57 (83.82)	1.00 [Reference]	137 (90.73)	1.00 [Reference]
CT+TT			11 (16.18)	0.83 (0.45, 1.54)	14 (9.27)	**0.49 (0.30, 0.79)**
Trend:				p=0.55		**p=0.003**
rs62023605#	16	*CLEC16A,*				
AA		*SOCS1*	39 (57.35)	1.00 [Reference]	71 (47.02)	1.00 [Reference]
AC^+CC^			29 (42.65)	1.55 (0.94, 2.54)	80 (52.98)	**1.78 (1.18, 2.68)**
Trend:				p=0.08		**p=0.006**

Abbreviations: SNPs: single-nucleotide polymorphisms; CDMS: Clinically definite MS.HR: hazard ratio.
* adjusted for age, sex, and study recruitment, before adjustment for multiple comparisons.
$ Provide nearest two genes for intergenic SNPs.
#The homozygous genotypes were combined with the heterozygous ones due to small numbers.
^ Risk genotype for cumulative genetic risk score (CGRS).
Results in bold denote statistically significant results (p<0.05).

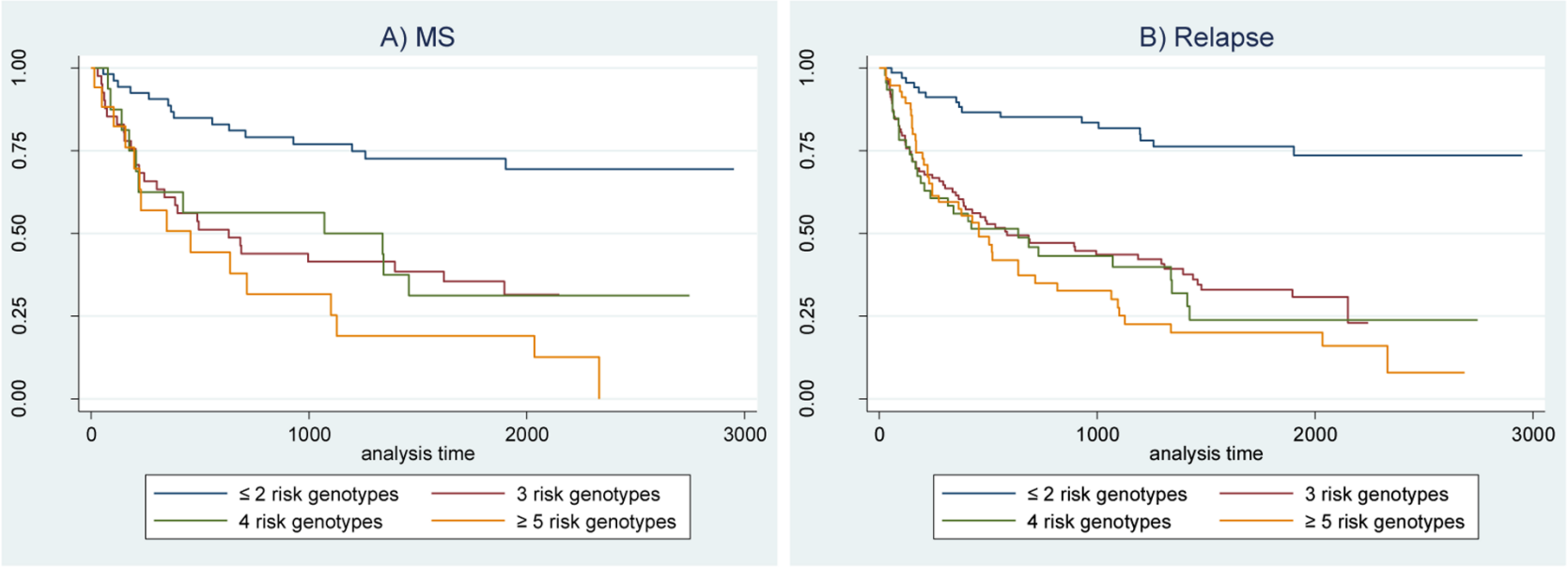
Combining the seven SNPs that predicted conversion to MS and/or relapse (Table 1) into a cumulative genetic risk score, we found evidence of a significant positive association of increasing number of risk genotypes and subsequent hazard of CDMS and relapse (Table 2, Figure 1). While the associations were not neatly dose-dependent for CDMS or relapse, these results suggest that an increasing number of risk genotypes is deleterious for subsequent disease activity.

Table 2 Cumulative risk of CDMS and relapse for the seven SNPs associated with conversion to MS and relapse*

	CDMS		p-value	Relapse		p-value
	Number of CDMS	HR (95% CI)		Number of relapses	HR (95% CI)	
≤ 2 risk genotypes#	15	1.00 [Reference]		16	1.00 [Reference]	
3 risk genotypes	27	**3.49 (1.76, 6.92)**	**3.45×10^{-4}**	64	**3.91 (2.12, 7.27)**	**1.65×10^{-5}**
4 risk genotypes	11	**3.35 (1.61, 6.98)**	**1.27×10^{-3}**	30	**4.51 (2.39, 8.53)**	**3.44×10^{-6}**
≥ 5 risk genotypes*	15	**5.98 (2.98, 12.01)**	**4.77×10^{-7}**	41	**6.07 (3.26, 11.28)**	**1.22×10^{-8}**
			$p_{trend}=1.41\times10^{-7}$			**$p_{trend}=9.87\times10^{-9}$**

None carried zero risk genotype, three participants carried one risk genotype for CDMS.
* One participant carried seven risk genotypes for CDMS, five participants carried six risk genotypes for CDMS.
Results presented are adjusted for age, sex, and study recruitment center (QLD, NSW, VIC, TAS).
Results in bold denote statistically significant results ($p<0.05$).

Figure 1: Kaplan-Meier survival plot for cumulative genetic risk score of conversion to MS (A) and relapse (B).



SNP Predictors of annualised change in disability

We identified seven non-HLA SNPs (Table 3) that were associated with ΔEDSS; no HLA SNPs significantly predicted disability progression. None of these SNPs showed any material association with conversion to MS or relapse.

Table 3 Seven top SNPs and their associations with annualized change in disability (ΔEDSS)

SNP	Chr	Gene[$]	ΔEDSS	
			Number of 5-year disability measures (%)	β (95%CI)
rs7588193[#]	2	*ZFP36L2 ,*		
AA		*HAAO*	74 (59.20)	0.27 (0.22, 0.32)
AG^+GG^			51 (40.80)	**+0.12 (0.04, 0.21)**
Trend:				**p=0.005**
rs842639	2	*FLJ16341*		
AA^			61 (48.80)	0.36 (0.30, 0.42)
AG			48 (38.40)	-0.03 (-0.12, 0.06)
GG			16 (12.80)	**-0.20 (-0.32, -0.08)**
Trend:				**p=0.007**
rs35967351	1	*SLAMF7*		
AA^			60 (48.00)	0.38 (0.32, 0.44)
AT			54 (43.20)	**-0.10 (-0.19, -0.02)**
TT			11 (8.80)	-0.14 (-0.29, 0.01)
Trend:				**p=0.011**
rs2283792	22	*MAPK1*		
CC^			27 (21.60)	0.43 (0.33, 0.53)
CA			64 (51.20)	**-0.12 (-0.24, -0.01)**
AA			34 (27.20)	**-0.17 (-0.29, -0.04)**
Trend:				**p=0.013**
rs3825568	14	*ZFP36L1*		
GG			33 (27.05)	0.26 (0.18, 0.34)
GA^			59 (48.36)	+0.06 (-0.04, 0.16)
AA^			30 (24.59)	**+0.15 (0.03, 0.27)**
Trend:				**p=0.016**
rs2546890	5	*LOC285626*		
AA			46 (36.80)	0.26 (0.19, 0.32)
AG^			52 (41.60)	+0.09 (-0.01, 0.19)
GG^			27 (21.60)	**+0.12 (0.01, 0.24)**
Trend:				**p=0.030**
rs8070345	17	*VMP1*		
GG			32 (25.60)	0.26 (0.18, 0.34)
GA^			61 (48.80)	+0.06 (-0.05, 0.16)
AA^			32 (25.60)	**+0.12 (0.0009, 0.24)**
Trend:				**p=0.047**

Abbreviations: SNPs: single-nucleotide polymorphisms; CDMS: Clinically definite MS.HR: hazard ratio; EDSS: Expanded Disability Status Scale. ΔEDSS: Annualised disability progression from FDE to 5-year review.

* adjusted for age, sex, study recruitment center and whether participants were having a relapse at the time of their 5-year disability measurement, before adjustment for multiple comparisons.

$ Provide nearest two genes for intergenic SNPs.
#The homozygous genotypes were combined with the heterozygous ones due to small numbers.
^ Risk genotype for cumulative genetic risk score (CGRS).
Disability results presented as geometric mean ΔEDSS (95% CI) for the reference group, while coefficient relative to reference (β (95% CI)) are presented for subsequent levels.
Results in boldface denote statistically significant results ($p<0.05$).

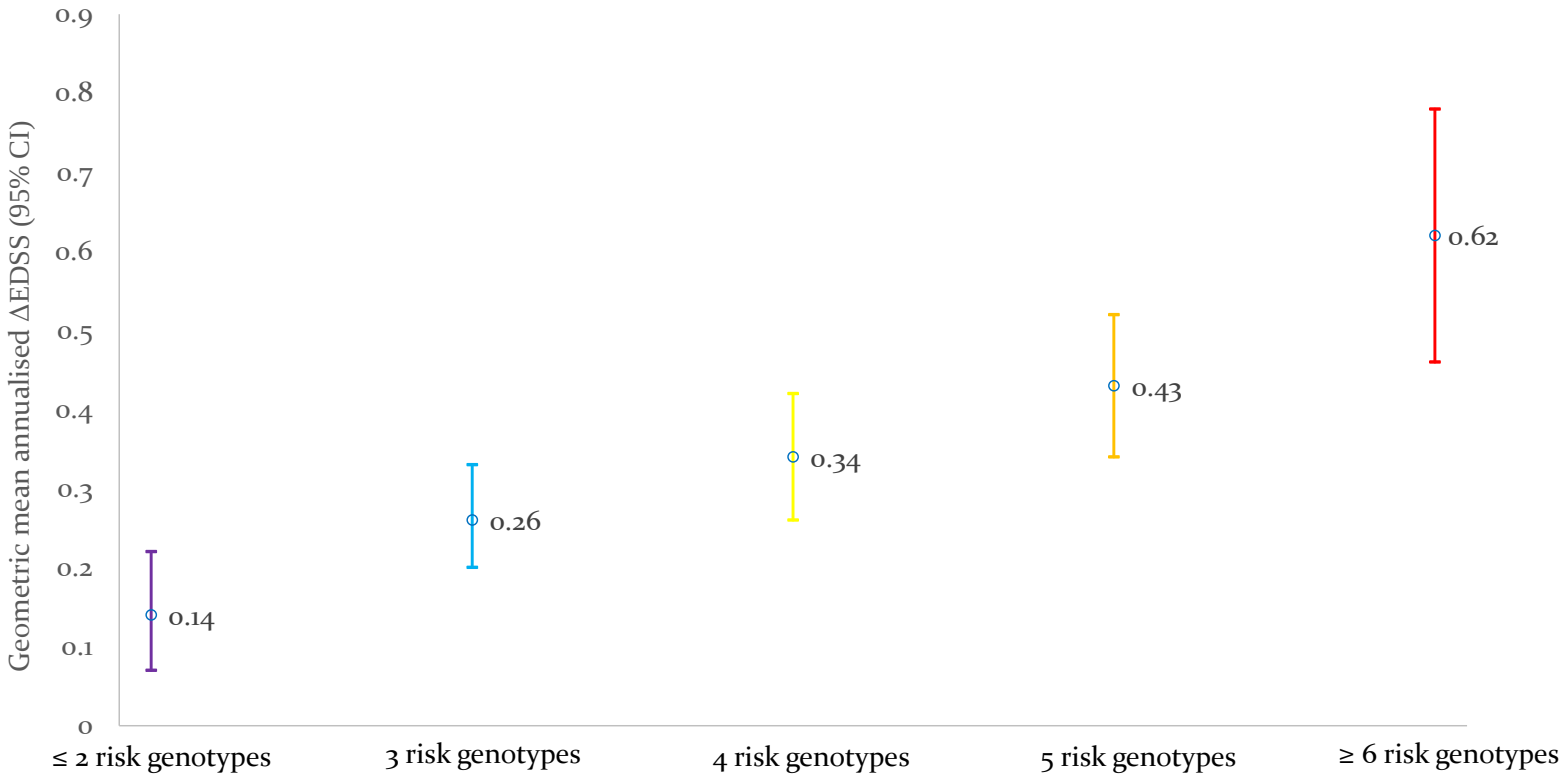
For the seven disability-associated SNPs, where the risk genotype was the genotype associated with an increase in EDSS and not necessarily the minor allele, we found a strong and significant dose-response (Table 4, Figure 2). For example compared to those with ≤2 risk genotypes, those with ≥6 risk genotypes had an annual disability progression rate of nearly 0.5 EDSS points greater, which over five years equates to 2.5 EDSS points. The CGRS model explained 32.7% of the variance in disability progression ($R^2=0.327$, $P_{trend}=1.53\times10^{-9}$).

Table 4 Cumulative risk of disability for the seven SNPs that predicted ΔEDSS

	Number of 5-year disability measures	β (95%CI)	p-value	R^2
≤ 2 risk genotypes[#]	21	0.14 (0.07, 0.22)		0.327
3 risk genotypes	39	**+0.12 (0.02, 0.22)**	**0.024**	
4 risk genotypes	29	**+0.20 (0.09, 0.31)**	**6.80×10^{-4}**	
5 risk genotypes	26	**+0.28 (0.17, 0.40)**	**5.14×10^{-6}**	
≥ 6 risk genotypes*	10	**+0.48 (0.30, 0.66)**	**8.36×10^{-8}**	
			$p_{trend}=1.53\times10^{-9}$	

No participants carried zero risk genotypes, five participants carried one risk genotypes, sixteen carried two risk genotypes.
* No participants carried seven risk genotypes, ten participants carried 6 risk genotypes.
p: adjusted for age, sex, and study recruitment centre; ΔEDSS: annualised change in EDSS.
Disability results presented as geometric mean ΔEDSS (95% CI) for the reference group, and coefficient relative to the reference (β (95% CI)) for subsequent levels.
Results in bold denote statistically significant results ($p<0.05$).

Figure 2: The line plot of cumulative genetic risk score predicting ΔEDSS. Results presented as geometric mean ΔEDSS and 95% CI.



HLA SNP predictors of MS clinical course

In addition to the 110 non-HLA MS-risk associated SNPs, we also examined the association of six HLA SNPs which have been associated with MS risk. These results show that only two of these were associated with the hazard of CDMS and relapse, but not disability, while the prototypical HLA MS-risk associated loci, *DRB1*1501*, was not significantly associated with any clinical outcomes in this study (Table 5).

Table 5 The association between three HLA SNPs and MS clinical course*

SNP	HLA allele or SNP	CDMS		Relapse		ΔEDSS	
		Number of CDMS (%)	HR (95% CI)	Number of relapse (%)	HR (95% CI)	Number of 5-year disability measures (%)	β(95%CI)
rs9266773[$]	*B*44:02*						
AA		60 (88.24)	1.00 [Reference]	142(94.04)	1.00 [Reference]	110 (88.00)	-0.86 (-0.93,-0.80)
AG		8 (11.76)	**0.44 (0.20, 0.97)**	9 (5.96)	**0.37 (0.19, 0.71)**	15 (12.00)	-0.04 (-0.19, 0.10)
Trend:			**p=0.043**		**p=0.003**		p=0.62
rs9277561[#]	rs9277565[T]						
AA		43 (64.18)	1.00 [Reference]	78 (52.35)	1.00 [Reference]	76 (61.79)	-0.89 (-0.96,-0.82)
AG+GG		24 (35.82)	1.08 (0.64, 1.84)	71 (47.65)	**1.59 (1.04, 2.43)**	47 (38.21)	+0.04 (-0.10, 0.18)
Trend:			p=0.77		**p=0.033**		p=0.54
rs3135391[#]	*DRB1*15:01*						
GG		25 (36.76)	1.00 [Reference]	50(33.11)	1.00 [Reference]	54 (43.20)	-0.80 (-0.94,-0.65)
GA+AA		43 (63.24)	1.63 (0.97, 2.73)	101(66.89)	1. 30 (0.82, 2.06)	71 (56.80)	-0.11 (-0.27, 0.05)
Trend:			p=0.067		p=0.26		p=0.12

Abbreviations: SNPs: single-nucleotide polymorphisms; HR: hazard ratio; EDSS: Expanded Disability Status Scale; ΔEDSS: Annualised disability progression from FDE to 5-year review.
* adjusted for age, sex, and study recruitment, before adjustment for multiple comparisons.
CDMS: Clinically definite MS, FDE: First demyelinating event.

$ No persons had the GG genotype.
#The homozygous genotypes were combined with the heterozygous ones due to small numbers.
Disability results presented as geometric mean ΔEDSS (95% CI) for the reference group, while coefficient relative to reference (β (95% CI)) are presented for subsequent levels.
Results in boldface denote significant results ($p<0.05$).

2.8 Discussion

Using a longitudinal cohort of participants with a first neurological presentation of symptoms suggestive of CNS demyelination, we investigated whether known MS susceptibility SNPs were associated with MS clinical course and disability progression in early disease. We found that several known MS risk-associated SNPs influenced MS clinical course, including 7 SNPs which predicted the hazard of CDMS and/or relapse and 7 other SNPs which predicted ΔEDSS. There is a greater chance that SNPS found in both the relapse and conversion to MS analyses are more likely to be real findings than those found in either analysis alone as these 2 measures (relapse and conversion to MS) measure the same parameter but in different ways. While none of these SNPs individually remained significant after adjusting for multiple comparisons, epidemiological supports such as dose-dependency and internal consistency between related clinical outcomes supported the validity of taking these SNPs forward to a CGRS assessment. The CGRS analysis showed that, in combination, a greater number of risk genotypes had a highly positive association with conversion to MS (HR 5.98 for ≥ 5 risk genotypes vs ≤ 2 risk genotypes), relapse (HR 6.07 for ≥ 5 risk genotypes vs ≤ 2 risk genotypes), and ΔEDSS where change in EDSS for those who had ≥ 6 risk genotypes was 0.48 EDSS points per year greater than reference.

Our CGRS model for disability progression explained 32.7% of the variance in MS disability progression within this dataset. These results suggest that these seven common variants in combination significantly contribute to disability progression of multiple sclerosis.

The risk variants detected were completely different between disability and CDMS/relapse. Hence, we hypothesised that CDMS/relapse and disability progression may be driven by different genetic pathways, with CDMS and relapse driven by central nervous system inflammation, whereas disability progression may be more driven by neurodegeneration.

Previous work has suggested that the two processes may be independent(23), although this is controversial. The lack of overlap between genetic variants that may drive conversion and relapse and those associated with disability progression is of great interest and may add support to the argument that these two processes may be independent and require different approaches to treatment.

One interesting observation in our study was that the effects on MS clinical course of the HLA SNPS that have such significant effects on MS risk was varied, with only HLA-B*44:02 (rs9266773) having a significant protective association with relapse and conversion to MS, the latter reaching statistical significance on permutation testing after correction for multiple testing. The MS risk allele of HLA DRB1*15:01 was not clearly associated with MS clinical course in this study, supporting findings from some but not all previous studies(2, 10). We did not find that *HLA DRB1*1501* was associated with MS clinical course and the *HLA B*44:02* proxy SNP was not significantly associated with any element of MS clinical course, thus supporting findings from other studies(2). It is reasonable to assume that, while some risk-associated SNPs are also associated with MS clinical course, others may only predict onset.

We have shown some overlap with our previous study in established MS that further validates this work. In particular, the MS risk SNP near the *RGS1* gene associated with the hazard of CDMS in the current analysis was significantly associated with subsequent relapse risk in our previous study(10).

Basing results only on statistical significance in a longitudinal MS study when looking at multiple genetic markers is difficult and requires large sample sizes. The major limitation of our study is the small sample size, particularly when this is further reduced by restriction to only those with genotyping data and those with initial bout-onset disease with onset close to the time of study entry. Therefore, in our study, we have also used several other epidemiological concepts to provide support for our results, including dose-dependency of

allelic effect, internal consistency between related outcome measures (CDMS and relapse), and external consistency of directionality with associations found previously, as well as cumulative genetic risk scores. All 7 SNPs that were associated with CDMS and relapse risk, had significant allele dose responses, and all effects were in the same direction for the hazard of CDMS and relapse and in the same direction as for MS risk in GWAS providing support for their significance. These seven SNPs may be near genes that have significant effects on MS clinical course and warrant further investigation.

A key strength of our study is its long follow-up, beginning at the first presentation of symptoms of disease and continuing for now 10 years from onset. This allows confidence that the clinical course parameters measured are accurate, particularly for disability progression. Large GWAS analyses, while benefiting from a large sample size allowing for the ability to adjust for multiple comparisons, are methodologically limited by their inability to do more than compare groups, or measure progression using cross-sectional measures, rather than using time-to-event prospective analyses of clinical course that we have used in the present study. In this study, we have utilised the study strengths of a prospective cohort study design and evaluated the known MS risk-associated SNPs as predictors of clinical course. In this fashion, we retain the methodological strengths of the study design, the accuracy of prospective clinical course monitoring, and the reduction of reverse causality, while not having the statistical limitations of trying to evaluate using a genome-wide approach. We have utilised this approach previously in our cohort of established MS (average disease duration 12 years). However, that study was undertaken in a cohort that experienced little disability progression over a mean follow-up of 2.3 years and was in a largely treated population with a low annual relapse rate. The present study makes use of a cohort followed essentially from symptom onset and who accordingly were not on disease-modifying therapy or yet suffering appreciable impacts of disease. As relapsing-remitting MS patients have a highly variable time interval between the first and the second episode of central nervous

system (CNS) demyelination which clinically or radiologically defines the onset of CDMS(24).Understanding the genetic determinants of this temporal window of disease clinical course is important as this could allow appropriate counselling, open new avenues for drug development, and allow better selection from the available treatment options. Even so, our results should be replicated in other longitudinal cohorts to allow greater confidence in their veracity.

In conclusion, our findings support an association between known MS risk genes and MS clinical course. These data support a role for genetic factors in MS progression and suggest that the genetic drivers of MS progression are polygenic. These results require validation in other cohorts, but, with replication these loci may serve as potential targets for further translational research.

Acknowledgements

The members of the AUSLONG Investigator Group are: Robyn M Lucas (National Centre for Epidemiology and Population Health, Canberra), Keith Dear (Duke Kunshan University, Kunshan, China), Anne-Louise Ponsonby and Terry Dwyer (Murdoch Childrens Research Institute, Melbourne, Australia), Ingrid van der Mei, Leigh Blizzard, Steve Simpson Jr and Bruce V Taylor (Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia), Simon Broadley (School of Medicine, Griffith University, Gold Coast Campus, Australia), Trevor Kilpatrick (Centre for Neurosciences, Department of Anatomy and Neuroscience, University of Melbourne, Melbourne, Australia). David Williams and Jeanette Lechner-Scott (University of Newcastle, Newcastle, Australia), Cameron Shaw and Caron Chapman (Barwon Health, Geelong, Australia), Alan Coulthard (University of Queensland, Brisbane, Australia) and Patricia Valery (QIMR Berghofer Medical Research Institute, Brisbane, Australia).

2.9 References

1. International Multiple Sclerosis Genetics C, Beecham AH, Patsopoulos NA, Xifara DK, Davis MF, Kemppinen A, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat Genet. 2013;45(11):1353-60.
2. International Multiple Sclerosis Genetics C, Wellcome Trust Case Control C, Sawcer S, Hellenthal G, Pirinen M, Spencer CC, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011;476(7359):214-9.
3. International Multiple Sclerosis Genetics C, International IBDGC, International IBDGCI. Class II HLA interactions modulate genetic risk for multiple sclerosis. Nat Genet. 2015.
4. Baranzini SE, Wang J, Gibson RA, Galwey N, Naegelin Y, Barkhof F, et al. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. Hum Mol Genet. 2009;18(4):767-78.
5. Martinelli-Boneschi F, Esposito F, Brambilla P, Lindstrom E, Lavorgna G, Stankovich J, et al. A genome-wide association study in progressive multiple sclerosis. Mult Scler. 2012;18(10):1384-94.
6. Patsopoulos NA, Barcellos LF, Hintzen RQ, Schaefer C, van Duijn CM, Noble JA, et al. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. PLoS Genet. 2013;9(11):e1003926.
7. Australia, New Zealand Multiple Sclerosis Genetics C. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. Nat Genet. 2009;41(7):824-8.
8. Jensen CJ, Stankovich J, Van der Walt A, Bahlo M, Taylor BV, van der Mei IA, et al. Multiple sclerosis susceptibility-associated SNPs do not influence disease severity measures in a cohort of Australian MS patients. PLoS One. 2010;5(4):e10003.
9. van der Walt A, Stankovich J, Bahlo M, Taylor BV, van der Mei IA, Foote SJ, et al. Apolipoprotein genotype does not influence MS severity, cognition, or brain atrophy. Neurology. 2009;73(13):1018-25.
10. Lin R, Taylor BV, Simpson S, Jr., Charlesworth J, Ponsonby AL, Pittas F, et al. Association between multiple sclerosis risk-associated SNPs and relapse and disability--a prospective cohort study. Mult Scler. 2014;20(3):313-21.
11. Ponsonby AL, Lucas RM, Dear K, van der Mei I, Taylor B, Chapman C, et al. The physical anthropometry, lifestyle habits and blood pressure of people presenting with a first clinical demyelinating event compared to controls: The Ausimmune study. Mult Scler. 2013;19(13):1717-25.
12. Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". Ann Neurol. 2005;58(6):840-6.
13. McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Ann Neurol. 2001;50(1):121-7.
14. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). Neurology. 1983;33(11):1444-.
15. Guo Y, He J, Zhao S, Wu H, Zhong X, Sheng Q, et al. Illumina human exome genotyping array clustering and quality control. Nat Protoc. 2014;9(11):2643-62.

16. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38(8):904-9.
17. Niino M, Sato S, Fukazawa T, Yoshimura S, Hisahara S, Matsushita T, et al. Latitude and HLA-DRB1 alleles independently affect the emergence of cerebrospinal fluid IgG abnormality in multiple sclerosis. Mult Scler. 2015.
18. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. Biometrika. 1981;68(2):373-79.
19. Gordi T, Khamis H. Simple solution to a common statistical problem: interpreting multiple tests. Clin Ther. 2004;26(5):780-6.
20. Lin R, Taylor BV, Simpson S, Jr., Charlesworth J, Ponsonby AL, Pittas F, et al. Novel modulating effects of PKC family genes on the relationship between serum vitamin D and relapse in multiple sclerosis. J Neurol Neurosurg Psychiatry. 2014;85(4):399-404.
21. Paynter NP, Chasman DI, Pare G, Buring JE, Cook NR, Miletich JP, et al. Association between a literature-based genetic risk score and cardiovascular events in women. JAMA. 2010;303(7):631-7.
22. Zhou XJ, Qi YY, Hou P, Lv JC, Shi SF, Liu LJ, et al. Cumulative effects of variants identified by genome-wide association studies in IgA nephropathy. Sci Rep. 2014;4:4904.
23. Confavreux C, Vukusic S, Moreau T, Adeleine P. Relapses and progression of disability in multiple sclerosis. N Engl J Med. 2000;343(20):1430-8.
24. Sadaka Y, Verhey LH, Shroff MM, Branson HM, Arnold DL, Narayanan S, et al. 2010 McDonald criteria for diagnosing pediatric multiple sclerosis. Ann Neurol. 2012;72(2):211-23.

Chapter 3. Conclusion

3.1 On the role of risk-associated genetic loci in modulating clinical course in multiple sclerosis.

In multiple sclerosis (MS), patients generally present with a first clinical episode of demyelination (FDE) which has an uncertain prognosis with few generalizable markers of subsequent disease severity. Following an FDE there are no clear genetic indicators of future conversion to MS (CDMS), risk of relapse, or rate of disability accumulation, hampering the effort to providing early intervention to those patients at high risk.

We therefore, developed an a priori hypothesis that 116 MS-risk associated genetic single nucleotide polymorphisms (SNPs) may also significantly affect MS clinical outcomes. Utilising a longitudinal cohort study of persons who have had a FDE, followed for 5-years, we found that:

- Seven non-HLA SNPs predicted relapse and/or CDMS, and seven other non-HLA SNPs predicted the annualised change in disability status (ΔEDSS, as measured by Expanded Disability Status Scale)
- We generated genetic risk score (CGRS) based on those identified risk SNPs, which significantly predicted CDMS and relapse in a significant, dose-dependent manner: patients having >5 risk genotypes had a 6-fold greater risk of CDMS and relapse compared to those with <2; those carrying ≥6 risk genotypes progressed at 0.48 EDSS points per year faster compared to those with <2, and the CGRS model explained 32% of the variance in disability.

For the first time, we demonstrated that MS genetic susceptibility SNPs predicted worse MS clinical course in three key metrics. These results, if replicated, may aid in developing prognostic algorithms in the early disease period in MS and provide further mechanistic insights. We believe these findings will be of benefit for MS patients.

The candidate-gene approach and GWAS have been the two major methods for genetic research. Linkage studies based on family data have been useful in examining the effect of specific genes associated with Mendelian (i.e. monogenic) diseases, but it has been less effective in identifying susceptibility genes of polygenic complex disease such as MS. Likewise, the candidate-gene approach has usually been unsuccessful in identifying novel loci in MS genetic research. Whereas GWAS is more useful and effective to search MS susceptibility genes. Thus, it enhances the chance to identify MS clinical course associated variants by utilising candidate gene which identified by GWAS in our longitudinal cohort study.

3.2 Major Contributions

My research's major contributions can be listed as follows:

- Using one of the largest longitudinal studies of early-MS clinical course, we present evidence that the cumulative effects MS susceptibility genes significantly modulate MS clinical course.
- These data strongly support a role for genetic factors in MS progression and strongly suggest that the genetic drivers of MS progression are polygenic.
- These loci may serve as potential targets for further translational research.

3.3 Future directions and challenge

The AusLong study cohort has huge environmental, genetic and clinical data, including conversion to clinically definite MS (CDMS) and occurrence of relapse derived from annual review and medical records, while EDSS was measured at 5-year review. Our research group has collected extant environmental and clinical data on these participants over the last 9-10 years. The 10-year review data collection will be finished soon, future analysis utilising these data can be listed as follows:

- Clinical application: Translational medicine - translating predictive factors to clinical application. For genetics, this means potentially utilising genetic testing of at-risk patients to identify their potential future risk of disease (e.g. offspring/relatives of existing cases) or gauging prognosis in persons presenting with symptoms suggestive of the disease.
- Biological mechanisms: having substantiated some MS-risk associated SNPs in MS clinical course, now it remains to identify the mode by which these manifest in disease and impact on disease severity. For some genes this is more evident, but SNPs in intergenic regions will require investigation.
- Identify other potential polymorphism targets, gene coding functions and epigenetic regulation: For genes, it means using rare variants (minor allele frequencies (MAF<5%)) exome chips in large sample size to find new rare polymorphism targets which associated with MS onset and clinical course. For gene coding experiment, it is necessary to identify how these risk variants impair or change the function of the resulting proteins and how these changed proteins influence MS onset and clinical course. With the development of epigenetic, it's possible to utilize methylation analysis such as genome-wide methylation profiling to identify and understand the functional mechanisms at work in MS (both MS onset and processes).

Up to the present, major progress has been made in identifying multiple sclerosis susceptibility genes. However, research in genetic modulators of clinical course and disease severity are comparatively limited. While this may partly reflect the inherent complexity of the condition, it may also reflect the need for different study designs that more effectively measure disease-related events and disability accumulation, while at the same time balancing against the logistical barriers attendant to this type of follow-up. Therefore, the new discovery will depend on recruiting large cohorts, while maintaining comprehensive databases of environmental/behavioural, genetic and clinical information, and potentially utilising novel genetic technologies and statistical models. For instance, using new exome chips which focus

on the rare variants and well-measured phenotypes such as CDMS, relapse, disability, and brain atrophy (identified by MRI). Further research into these MS susceptibility genes may lead to a greater understanding of multiple sclerosis susceptibility & pathogenesis, and to a more personalised treatment approach.

Appendix 1. List of 110 known non-HLA MS risk variants

Chr	rsID	Position	RAF	RA	OR (95% CI)	P-value	Gene	Function
1	rs3748817	2525665	0.64	A	1.14 (1.10-1.18)	1.33E-12	*MMEL1*	intronic
1	rs3007421	6530189	0.12	A	1.12 (1.07-1.18)	9.61E-07	*PLEKHG5*	intronic
1	rs12087340	85746993	0.09	A	1.22 (1.15-1.29)	5.13E-12	*BCL10 (dist=4406), DDAH1 (dist=37175)*	intergenic
1	rs11587876	85915183	0.79	A	1.12 (1.07-1.17)	8.40E-08	*DDAH1*	intronic
1	rs41286801	92975464	0.14	A	1.20 (1.15-1.25)	7.92E-16	*EVI5*	UTR3
1	rs7552544	101240893	0.56	A	1.08 (1.05-1.12)	3.67E-06	*VCAM1 (dist=36292), EXTL2 (dist=97035)*	intergenic
1	rs11581062	101407519	0.29	G	1.05 (1.01-1.09)	1.20E-02	*VCAM1*	intronic
1	rs6677309	117080166	0.88	A	1.34 (1.27-1.41)	1.45E-28	*CD58*	intronic
1	rs666930	120258970	0.53	G	1.09 (1.06-1.13)	7.49E-08	*PHGDH*	intronic
1	rs2050568	157770241	0.53	G	1.08 (1.05-1.12)	1.33E-06	*FCRL1*	intronic
1	rs35967351	160711804	0.67	A	1.09 (1.05-1.13)	1.70E-06	*SLAMF7*	intronic
1	rs1359062	192541472	0.82	C	1.18 (1.13-1.23)	1.84E-13	*RGS21 (dist=205058), RGS1 (dist=3385)*	intergenic
1	rs55838263	200874728	0.71	A	1.12 (1.08-1.17)	1.41E-09	*C1orf106*	intronic
2	rs4665719	25017860	0.25	G	1.09 (1.05-1.13)	6.80E-06	*CENPO*	intronic
2	rs2163226	43361256	0.71	A	1.10 (1.07-1.15)	7.02E-08	*HAAO (dist=341505), ZFP36L2 (dist=88285)*	intergenic

2	rs842639	61095245	0.65	A	1.11 (1.08-1.15)	1.70E-09	*FLJ16341*	ncRNA_intronic
2	rs7595717	68587477	0.26	A	1.10 (1.06-1.14)	3.29E-07	*CNRIP1 (dist=40294), PLEK (dist=4845)*	intergenic
2	rs17174870	112665201	0.76	G	1.03 (1.00-1.07)	8.84E-02	*MERTK*	intronic
2	rs9967792	191974435	0.62	G	1.11 (1.07-1.15)	1.80E-09	*STAT4*	intronic
2	rs9989735	231115454	0.18	C	1.17 (1.12-1.22)	7.84E-14	*SP140*	intronic
3	rs11719975	18785585	0.27	C	1.09 (1.05-1.13)	5.39E-06	*SATB1 (dist=305320), KCNH8 (dist=404432)*	intergenic
3	rs2371108	27757018	0.38	A	1.08 (1.05-1.12)	2.06E-06	*EOMES*	downstream
3	rs1813375	28078571	0.47	A	1.15 (1.12-1.19)	5.75E-18	*EOMES (dist=314786), CMC1 (dist=204553)*	intergenic
3	rs4679081	33013483	0.52	G	1.08 (1.04-1.11)	1.20E-05	*CCR4 (dist=17080), GLB1 (dist=24617)*	intergenic
3	rs9828629	71530346	0.62	G	1.08 (1.05-1.12)	5.49E-06	*FOXP1*	intronic
3	rs2028597	105558837	0.92	G	1.04 (0.98-1.11)	1.79E-01	*CBLB*	intronic
3	rs1131265	119222456	0.80	C	1.19 (1.14-1.24)	1.97E-15	*TIMMDC1*	exonic
3	rs1920296	121543577	0.64	C	1.14 (1.11-1.18)	6.75E-15	*IQCB1*	intronic
3	rs2255214	121770539	0.52	C	1.11 (1.08-1.15)	1.72E-10	*ILDR1 (dist=29412), CD86 (dist=3670)*	intergenic
3	rs9282641	121796768	0.92	G	1.12 (1.05-1.19)	5.86E-04	*CD86*	UTR5
3	rs1014486	159691112	0.43	G	1.11 (1.07-1.14)	1.16E-09	*IQCJ-SCHIP1 (dist=75957), IL12A (dist=15511)*	intergenic
4	rs7665090	103551603	0.52	G	1.08 (1.05-1.12)	2.41E-06	*NFKB1 (dist=13144), MANBA (dist=1040)*	intergenic

4	rs2726518	106173199	0.55	C	1.09 (1.05-1.13)	1.23E-05	*TET2*	intronic
5	rs6881706	35879156	0.72	C	1.12 (1.08-1.16)	4.87E-09	*IL7R (dist=2233), CAPSL (dist=25242)*	intergenic
5	rs6880778	40399096	0.60	G	1.10 (1.06-1.14)	1.70E-08	*DAB2 (dist=973761), PTGER4 (dist=280936)*	intergenic
5	rs71624119	55440730	0.76	G	1.12 (1.08-1.17)	2.70E-09	*ANKRD55*	intronic
5	rs756699	133446575	0.87	A	1.12 (1.07-1.18)	2.97E-06	*VDAC1 (dist=105751), TCF7 (dist=3827)*	intergenic
5	none	141506564	0.61	C	1.07 (1.04-1.11)	5.96E-05	*NDFIP1*	intronic
5	rs2546890	158759900	0.52	A	1.06 (1.02-1.09)	6.59E-04	*LOC285626*	ncRNA_exonic
5	rs4976646	176788570	0.34	G	1.13 (1.09-1.17)	1.04E-12	*RGS14*	intronic
6	rs17119	14719496	0.81	A	1.11 (1.06-1.15)	1.91E-06	*CD83 (dist=582348), JARID2 (dist=526710)*	intergenic
6	rs941816	36375304	0.18	G	1.13 (1.08-1.18)	4.47E-09	*PXT1*	intronic
6	rs72928038	90976768	0.17	A	1.11 (1.07-1.16)	7.63E-07	*BACH2*	intronic
6	rs802734	128278798	0.69	A	1.03 (0.99-1.06)	1.58E-01	*THEMIS (dist=39022), PTPRK (dist=11126)*	intergenic
6	rs11154801	135739355	0.37	A	1.11 (1.07-1.15)	2.35E-09	*AHI1*	intronic
6	rs17066096	137452908	0.23	G	1.14 (1.10-1.18)	5.91E-12	*IL20RA (dist=86610), IL22RA2 (dist=12049)*	intergenic
6	rs7769192	137962655	0.55	G	1.08 (1.04-1.12)	1.30E-05	*OLIG3 (dist=147124), TNFAIP3 (dist=225670)*	intergenic

Appendix 1. List of 110 known non-HLA MS rick variants

6	rs67297943	138244816	0.78	A	1.12 (1.07-1.16)	4.83E-08	*TNFAIP3 (dist=40367), PERP (dist=164826)*	intergenic
6	rs212405	159470559	0.62	T	1.15 (1.11-1.19)	1.43E-15	*TAGAP (dist=4375), FNDC1 (dist=119870)*	intergenic
7	rs1843938	3113034	0.44	A	1.08 (1.05-1.12)	2.21E-06	*CARD11 (dist=29525), SDK1 (dist=228046)*	intergenic
7	rs706015	27014988	0.18	C	1.14 (1.09-1.19)	1.29E-09	*SKAP2 (dist=110647), HOXA1 (dist=117626)*	intergenic
7	rs917116	28172739	0.20	C	1.12 (1.07-1.16)	2.07E-08	*JAZF1*	intronic
7	rs60600003	37382465	0.10	C	1.16 (1.10-1.22)	2.53E-08	*ELMO1*	intronic
7	rs201847125	50325567	0.70	G	1.11 (1.07-1.15)	2.91E-08	*C7orf72 (dist=126715), IKZF1 (dist=18811)*	intergenic
7	rs354033	149289464	0.74	G	1.03 (1.00-1.07)	7.70E-02	*ZNF767*	ncRNA_intronic
8	rs1021156	79575804	0.24	A	1.12 (1.08-1.16)	5.60E-10	*PKIA (dist=58302), ZC2HC1A (dist=2478)*	intergenic
8	rs2456449	128192981	0.36	G	1.10 (1.06-1.14)	2.21E-08	*PCAT1 (dist=159722), POU5F1B (dist=234876)*	intergenic
8	rs4410871	128815029	0.72	G	1.12 (1.08-1.16)	1.98E-09	*MIR1204(dist=6755),PVT1(dist=87845)*	intergenic
8	rs759648	129158945	0.31	C	1.09 (1.05-1.13)	2.82E-06	*PVT1 (dist=45446), MIR1208 (dist=3417)*	intergenic
9	rs2150702	5893861	0.49	G	1.16 (1.10-1.22)	3.30E-08	*MLANA*	intronic
10	rs2104286	6099045	0.72	A	1.21 (1.16-1.26)	7.61E-23	*IL2RA*	intronic

10	rs793108	31415106	0.50	A	1.09 (1.06-1.13)	5.61E-08	*ZNF438 (dist=94240), ZEB1-AS1 (dist=190351)*	intergenic
10	rs2688608	75658349	0.55	A	1.07 (1.03-1.10)	6.37E-05	*CAMK2G (dist=24000), C10orf55 (dist=11378)*	intergenic
10	rs1782645	81048611	0.43	A	1.09 (1.05-1.13)	4.30E-07	*ZMIZ1*	intronic
10	rs7923837	94481917	0.61	G	1.11 (1.07-1.14)	4.58E-09	*HHEX (dist=26509), EXOC6 (dist=112553)*	intergenic
11	rs7120737	47702395	0.15	G	1.13 (1.08-1.18)	7.61E-08	*AGBL2*	intronic
11	rs34383631	60793330	0.40	A	1.11 (1.07-1.15)	5.69E-10	*CD6 (dist=5482), CD5 (dist=76600)*	intergenic
11	rs694739	64097233	0.62	A	1.08 (1.04-1.11)	1.30E-05	*PRDX5 (dist=7938), CCDC88B (dist=10457)*	intergenic
11	rs533646	118566746	0.68	G	1.10 (1.06-1.14)	3.60E-07	*TREH (dist=16365), DDX6 (dist=51727)*	intergenic
11	rs9736016	118724894	0.63	T	1.10 (1.07-1.14)	2.20E-08	*DDX6 (dist=62922), CXCR5 (dist=29581)*	intergenic
11	rs523604	118755738	0.53	A	1.09 (1.05-1.13)	2.50E-07	*CXCR5*	intronic
12	rs1800693	6440009	0.40	G	1.14 (1.11-1.18)	6.92E-16	*TNFRSF1A*	intronic
12	rs12296430	6503500	0.19	C	1.14 (1.09-1.18)	3.62E-10	*LTBR (dist=2768), CD27-AS1 (dist=44667)*	intergenic
12	rs11052877	9905690	0.36	G	1.10 (1.07-1.14)	5.37E-09	*CD69*	UTR3
12	rs201202118	58182062	0.67	A	1.14 (1.10-1.18)	7.40E-13	*TSFM*	intronic
12	rs7132277	123593382	0.19	A	1.10 (1.06-1.15)	1.88E-06	*PITPNM2*	intronic
13	rs4772201	100086259	0.82	A	1.12 (1.07-1.17)	1.67E-07	*MIR548AN (dist=27705), TM9SF2 (dist=67469)*	intergenic

14	rs2236262	69261472	0.50	A	1.08 (1.04-1.11)	1.16E-05	*ZFP36L1*	intronic
14	rs4903324	75961511	0.19	A	1.10 (1.05-1.14)	9.62E-06	*JDP2 (dist=22107), BATF (dist=27273)*	intergenic
14	rs74796499	88432328	0.95	C	1.31 (1.21-1.42)	8.47E-11	*GALC*	intronic
14	rs12148050	103263788	0.35	A	1.08 (1.04-1.11)	1.47E-05	*TRAF3*	intronic
15	rs59772922	79207466	0.83	A	1.11 (1.06-1.15)	4.02E-06	*MORF4L1 (dist=17385), CTSH (dist=6626)*	intergenic
15	rs8042861	90977333	0.44	A	1.08 (1.05-1.12)	9.80E-07	*IQGAP1*	intronic
16	rs2744148	1073552	0.18	G	1.09 (1.04-1.13)	1.02E-04	*SOX8 (dist=36573), SSTR5-AS1 (dist=40530)*	intergenic
16	rs12927355	11194771	0.68	G	1.21 (1.17-1.26)	8.19E-27	*CLEC16A*	intronic
16	rs4780346	11288806	0.23	A	1.09 (1.05-1.13)	6.80E-06	*CLEC16A (dist=12760), SOCS1 (dist=59468)*	intergenic
16	rs6498184	11435990	0.81	G	1.15 (1.10-1.21)	2.07E-10	*PRM1 (dist=60798), RMI2 (dist=3321)*	intergenic
16	rs7204270	30156963	0.50	G	1.09 (1.06-1.13)	9.32E-08	*MAPK3 (dist=22333), CORO1A (dist=37768)*	intergenic
16	rs1886700	68685905	0.14	A	1.11 (1.06-1.16)	8.76E-06	*CDH3*	intronic
16	rs12149527	79110596	0.47	A	1.08 (1.05-1.12)	1.74E-06	*WWOX*	intronic
16	rs7196953	79649394	0.29	A	1.08 (1.04-1.12)	2.65E-05	*MAF (dist=14772), DYNLRB2 (dist=925460)*	intergenic
16	rs35929052	85994484	0.89	G	1.14 (1.09-1.20)	3.32E-07	*IRF8 (dist=38273), LOC146513 (dist=325553)*	intergenic

17	rs12946510	37912377	0.47	A	1.08 (1.04-1.11)	8.51E-06	*GRB7 (dist=8839), IKZF3 (dist=1591)*	intergenic
17	rs4796791	40530763	0.36	A	1.10 (1.06-1.14)	1.81E-08	*STAT3*	intronic
17	rs4794058	45597098	0.50	A	1.07 (1.04-1.11)	1.63E-05	*MRPL45P2 (dist=27112), NPEPPS (dist=11346)*	intergenic
17	rs8070345	57816757	0.45	A	1.14 (1.11-1.18)	5.43E-16	*VMP1*	intronic
18	rs7238078	56384192	0.77	A	1.05 (1.02-1.10)	6.29E-03	*MALT1*	intronic
19	rs1077667	6668972	0.79	G	1.16 (1.12-1.21)	3.54E-13	*TNFSF14*	intronic
19	rs34536443	10463118	0.95	C	1.28 (1.18-1.40)	1.24E-08	*TYK2*	exonic
19	rs2288904	10742170	0.77	G	1.14 (1.09-1.19)	9.57E-10	*SLC44A2*	exonic
19	rs1870071	16505106	0.29	G	1.12 (1.08-1.16)	5.68E-10	*EPS15L1*	intronic
19	rs11554159	18285944	0.73	G	1.15 (1.11-1.20)	2.58E-13	*IFI30*	exonic
19	rs8107548	49870643	0.25	G	1.09 (1.05-1.13)	1.98E-06	*DKKL1*	intronic
20	rs4810485	44747947	0.25	A	1.08 (1.04-1.12)	1.78E-05	*CD40*	intronic
20	rs17785991	48438761	0.35	A	1.09 (1.05-1.13)	6.42E-07	*SLC9A8*	intronic
20	rs2248359	52791518	0.60	G	1.07 (1.03-1.10)	9.81E-05	*CYP24A1 (dist=1002), PFDN4 (dist=32984)*	intergenic
20	rs2256814	62373983	0.19	A	1.11 (1.07-1.16)	8.34E-07	*SLC2A4RG*	intronic
20	rs6062314	62409713	0.92	A	1.10 (1.03-1.16)	3.87E-03	*ZBTB46*	intronic
22	rs2283792	22131125	0.51	C	1.08 (1.05-1.12)	1.14E-06	*MAPK1*	intronic

Appendix 1. List of 110 known non-HLA MS rick variants

22	rs470119	50966914	0.39	A	1.07 (1.03-1.10)	1.51E-04	*TYMP*	intronic

Appendix 1A. Paper in Chapter 1

The potential role of genetic factors in multiple sclerosis onset and clinical course: a narrative review

Gongbu Pan, MS; Steve Simpson, Jr., PhD; Ingrid van der Mei, PhD; Jac C. Charlesworth, PhD;Yuan Zhou, MS; Feitong Wu, MS; Bruce Taylor, MD

Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia.

Abstract

Since multiple sclerosis (MS) was first identified as a complex autoimmune neurological disorder in the 19th century, a substantial body of research has been conducted to explore its aetiology and clinical course. Although many environmental and genetic factors have been associated with MS risk, substantiating their roles in its clinical course is comparatively more complex, highlighting an urgent need for further investigations. This review summarises findings from currently available research on MS clinical course and tries to give some future research directions.

1. Introduction

Multiple sclerosis (MS) is a chronic and immune-mediated central nervous system (CNS) condition characterized by inflammation and demyelination (1, 2), with typical onset between 20 and 40 years old(3). It is estimated that 2.5 million people are currently affected by MS worldwide, with a high prevalence (1/1,000) in Caucasian populations and with a high female/male sex ratio (between 1.5 and 2.5)(4, 5). Although the aetiology of this disease has not yet been fully understood, many MS genetic studies were established to get a well understanding of what associated with MS onset and clinical course. Genetic factors not only have been shown to play a crucial role in MS risk but also influenced MS clinical course. However, little is known about the role of genetic factors in MS clinical course, including conversion to clinically definite MS (CDMS), occurrence and frequency of relapse, and disability progression (6, 7). Genetic factors which modulate MS onset and clinical course, including the role of Human leukocyte antigen (HLA) and non-HLA variants, gene-environmental interactions.

In the last two decades, a total of 110 non-HLA region genetic loci and several HLA genetic loci associated with MS susceptibility have been identified by genome-wide

association studies (GWAS)(8). However, large GWAS have not found any associations between these MS susceptibility genes and MS clinical course. Consequently, the potential role of genetic factors in modulating MS clinical course remains unclear. In this review, we aim to summarise the findings of MS epidemiological and genetic studies and highlight the potential role of both environmental and genetic factors in MS onset and progression. Moreover, we outline and discuss challenges in this research, and the current gaps in our knowledge, including biological & clinical implications of genetic risk factors, potential clinical phenotype, and epigenetics that might be considered.

2. Genetic research method

Linkage and candidate gene method

Linkage analysis is based on the findings that genes located physically close together on a chromosome during the meiosis phase (9). Candidate gene methods are based on the knowledge of biological. During the linkage and candidate method period. The main findings of this period are HLA II arm area associated with MS onset which confirmed by linkage and candidate gene approach. Based on these findings, linkage and candidate gene approach were used to estimate the effect of substantial MS risk genes. Linkage analysis is a powerful tool based on the findings that genes located physically close together on a chromosome during the meiosis phase(9). Linkage analysis became the primary mode of statistical genetic mapping of Mendelian and complex diseases with familial aggregation in early diseases genetic research. Linkage study based on family data has been useful to examine the effect of specific genes associated with Mendelian (i.e., monogenic) diseases, but it has been less effective in identifying susceptibility genes for polygenic complex disease such as MS. In 2007, the International Multiple Sclerosis Genetics Consortium (IMSGC) established the largest non-parametric genome-wide linkage screen in 931 MS family trios. The important result for this study was identifying the significant linkage within the HLA locus, with a peak in the regional logarithm of the odds (LOD) score of 11.7, while none of the loci outside the HLA locus reached statistical significance. Then, lots of studies investigating genetic factors involved in MS onset, few susceptibility genes were identified. The *HLA-DRB1*15:01* allele had the greatest effect on risk of that region area. Generally, given the complexity of MS, and the potential for gene-environment interaction, linkage and candidate gene studies have not been successful

at finding susceptibility loci outside the MHC region in MS. Partly this is because these studies are limited by small sample size, as well as the small number of candidate variants from previous analyses. In the latter half of the 20 century, linkage and candidate analysis was the primary method used for the genetic research of Mendelian and complex diseases, which made a large contribution to disease genetic research.

Genome-wide association studies (GWAS)

Because the limitation of linkage and candidate gene method, genome wide association study (GWAS) testing millions of genetic variation marker, the single nucleotide polymorphism (SNP), have become the favored genetic mapping method of complex disease (10). To increase efficiency and cost, GWAS typically use a multistage design. Although GWAS have discovered many SNPs predicting a number of conditions (11), some major challenges have been raised.

1) **Ethnicity or racial mix**, referred to as population stratification, is a particular kind of confounding. People from different ancestries have a different allele frequency in some SNPs(12). GWAS can be confounded by population stratification which can lead to an increasing number of false positive results. The way to correct population stratification bias is to measure this confounder and adjust for it in regression models, or to stratify the case and controls.

2) **Multiple comparisons adjustment.** Statistical significance is usually defined as p=0.05 in classic epidemiological studies. However, for GWAS, the significance threshold must be adjusted to account for the substantial number of tests for SNPs. Bonferroni correction is widely used for multiple comparison adjustment. The Bonferroni method sets GWAS statistical significance at p=0.05/N. However, this method is sometimes considered to be too strict a criterion because some tests are not independent. In GWAS, many SNPs are in linkage disequilibrium (LD), which means they are correlated. To deal with this, the high number of test needs to be replaced. There are two common approaches to deal with this problem: 1) to calculate the true number of independent tests(13); 2) to use the false discovery rate (FDR) method to adjust the FDR for an independent test, which has been shown to be effective at improving statistical power (14). These methods suggest a genome

significant threshold between 10^{-6} and 10^{-8}, but the classic GWAS threshold of 10^{-8} remains widely used.

The preferable types of studies for MS onset and clinical course research.

Cross sectional study vs cohort study: a short photo in MS onset

Cross sectional studies are done to investigate whether people have MS onset or MS clinical course at a particular time. Cross sectional study looks like a snapshot for MS patient. GWAS used cross sectional case-control study design to investigate the relationship between the genetic factors and MS onset. Case-control studies (Genome-wide association study) look backwards. Beginning from an outcome, such as MS onset, this type of study works backward in time for genetic factors that might have caused the MS. Case-control studies are useful for MS onset that is rare, harmful and spends a long time on development, such as MS. There were many biases cannot avoid in the case-control study. For instance, recall bias, survival bias and surveillance bias etc(15, 16). Additionally, misclassification bias (the wrong diagnose of MS) is a problem in the case-control study that relies on a short time diagnosing.

In recent years, data mining have been used to test big data that proceeds without explicit hypotheses. All conclusions just based on the P values of statistical significance. Studies often use convenience samples without think of selection bias and other confounder factors can distort exposure-outcome associations. In addition, the MS clinical course (how the disease behaves over time) and long-term outcome of MS differs from patient to patient. By contrast, longitudinal cohort study such as Ausimmune/AusLong study which followed the progression of MS was more useful than case control study in MS genetic clinical course research(17). A longitudinal population based cohort study such as Ausimmune/AusLong study approaches to the study of the relationship between genetic and MS clinical course. These studies have specific hypotheses that inform MS clinical data collection and use classical statistic model to assess pre-specified genetic factors and MS clinical course phenotype. Cohort studies gave a logical sequence from genetic factors to MS clinical course. Hence, these longitudinal cohort study is easier to understand than a cross sectional case-control study. For Ausimmune/AusLong study, participants with a clinical diagnosis of a first demyelination episode which indicating a high risk of developing MS need to be recruited. Environmental, behavioural, genetic and clinical data,

including clinically isolated syndrome (CIS), conversion to clinically definite MS and occurrence of relapse, disability measured by EDSS and MSSS , Corpus callosum atrophy (CCA) need to be collected. Based on these data, the longitudinal cohort study can calculate a true odds ratio, hazard ratio, and confidence interval. However, each event (phenotype) for MS clinical takes many years to develop, this type of study design have a slow yield outcome and spend a lot of money.According to the small sample size of the most longitudinal study, P-value is hard to remain significant after multiple test. So several other epidemiological concepts should be done to support the results, including dose-dependency of allelic effect, internal consistency between related outcome measures (CDMS and relapse), and external consistency of directionality with associations found previously, as well as cumulative genetic risk scores.

3. The role of genetic factors in MS onset and clinical course

The role of HLA variants in MS onset

Of the genetic loci implicated in MS onset, none has been as consistent in magnitude and significance as the HLA region. The HLA region of the genome encodes the major histocompatibility complex (MHC), a key element of adaptive immunity, and spans a region of about 4,000Kb located on the short arm of chromosome 6 at band position 6p21.3. It is the most complex genetic polymorphism system of the human genome. The MHC region is divided into three major regions: HLA class I, encoding the MHC I proteins; class II, encoding the MHC II proteins; and class III, which are partly of unknown function but appear to play some role in the general immune function and regulation(18). The HLA class I and class II region are the most polymorphic region in human DNA sequences, leading to a great variety of genotypes, this a valuable component of adaptive immune recognition of the diverse number of epitopes to be recognised. The class I region contains *HLA-A*, *HLA-B* and *HLA-C* genes, while the class II region contains *HLA-DP*, *HLA-DQ* and *HLA-DR* genes. Variants in this regions have been associated with most inflammatory and autoimmune disease such as multiple sclerosis. Bertrams and colleagues identified that *HLA-A3* and *HLA-B7* were associated with MS onset(19). Subsequent study has confirmed this finding(20). Altogether, HLA variants have explained 10.5% of the heritability of MS(21). Of the HLA loci, most studies have shown *the HLA-DRB1*1501* locus to be the most strongly associated with MS risk (OR=3.1,

$P_{combined}<10^{-320}$)(22-24), connoting a three-fold greater risk of MS(25). Several other HLA alleles and haplotypes have also been identified to be associated with MS onset, but these appear to vary in their association with risk by ancestry and ethnicity. Stankovich and colleagues investigated HLA associations in 1,230 MS cases in Australia, finding not only *HLA-DRB1*1501* but also *HLA-DRB1*03* to be associated with MS risk (26). In Turkey and the Canary Islands, *HLA-DRB1*04* was associated with the risk of MS(27), while in Sardinia, *HLA-DRB1*0301* and *HLA-DRB1*0401* were identified as a genetic risk factor(27). MS patients in Canada and Sweden were associated with *HLA-DRB1*17*(28). *HLA-DRB1*1501* and *HLA-DQB1*0602* have an effect on in persons of African descent(23). Finally, even though Asian MS patients typically present as optico-spinal MS (OS-MS), in contrast to that seen in European-descent populations, MS patients in Japan also were associated with *HLA-DRB1*1501*(29); and the incidence of MS patients with *HLA-DRB1*0405* was significantly increased(30). Field and colleagues identified a relationship between MS and *HLA DRB1*15:01*, **0301*, **0401*, **1303*, *HLA-A*0201* and *HLA-DPB1*0301* in 1,618 MS case and *3,413* controls of European descent(31). A SNP in an intron of *HLA-DPB1*, rs3135021, was also associated with MS in African Americans(24). In the same year, a Polish group that a polymorphism in the *HLA-G* gene was associated with MS susceptibility (32). In 2012, Alcina and colleagues reported that the A allele of SNP rs3135388 in *DRB1*1501* was associated with *DQB1, DRB1* and *DRB5* genes' high expression in a Caucasian compared to G allele (33). These data demonstrate a significant role for the *DRB1*1501* locus in MS onset. Not only are these loci associated with MS onset, but also there is some evidence showing that deleterious HLA genotypes like *DRB1*1501* were positively associated with a marker of disease, oligoclonal bands(OB), while protective variants like *DRB1*0405* were negatively associated with OB(34). These findings are useful in that they provide internal consistency validation to HLA being associated with MS. Another study reported that the late onset MS was significantly associated with *HLA-DRB1*0801* gene(35). In 2009, Rama and other studies have found that *HLA-DR15*1501* have a significant effect on the age of onset(AO) of MS among Caucasians and African-Americans (36).

As shown by Chao and colleagues, women carry more MS risk allele *HLA-DRB1*15* than men. The prevalence of MS among women is approximately twice as in

men(37). Even sharing of the same genes and environmental risks, the concordance rate for the development of MS between identical twins is only 6-30%(38-41). These suggest that there are other mechanisms behind the scenario.

HLA & MS clinical course

The polymorphisms of HLA (especially *HLA-DRB1*) were highly associated with MS onset. However, little evidence showed that HLA genes were associated MS clinical course. One Australia study with 1230 MS cases and 1210 controls find both the *DRB1*04* and *DRB1*01/DRB1*15* genotype protected against PPMS (26). In 2010, another Australian study identified that *HLA-DRB1*1501* was correlated with MS severity as measured by MSSS; and *HLA-DRB1*1201* allele was correlated with less severe disease (42). Silva A.M et al did a case and control study with 248 patients and 282 healthy controls and reported that, in Portuguese, *HLA-DRB1*15* allele was a risk factor for MS and *HLA-DRB1*15* patients cause a longer time to reach a high EDSS (EDSS=3 or EDSS=6).(43). This result was in line with Weatherby's study (44). However, another study has reported a contradictory result that the *DRB1*01/04* and *DRB1*15/15* cause a shorter time to reach a high EDSS (EDSS=6) (45). In 2003, Barcellos et al found that there is a worse MS disease course in individuals with *HLA-DR2*(46). However, this result was not confirmed by subsequent study with a large sample size in the same group(47). In northeast Italy, *HLA-DR13* was associated with "benign" multiple sclerosis(48).But whether it can be taken as a prognostic factor need long time clinical follow up and large sample size. In west Australia, *HLA-DRB1*1501* was the strongest risk variants in RRMS and PPMS(49). The *DRB1*1501* also cause a worse disease course (50). Among African American MS patients, both *DRB1*1501* and *DRB1*1503* did not contribute to the disease severity, thus, MS severity was found to be associated with another gene or genes within HLA loci (51).

In short, there were few results about the relationship between HLA variants and MS clinical course because the case control study was widely designed to identify the relationship between exposures and outcome. However, MS relapse and disability progression last a long time period which can be followed by longitudinal cohort study. Thus, based on above results, it is reasonable to assume that the HLA variants contribute to MS onset and initial triggering mechanisms rather than modulating MS

clinical course. The main contribution of HLA variants is the influences of MS onset.

The role of non-HLA variants in MS onset & progression

During the last few decades, utilising candidate-gene, linkage studies and genome-wide association study (GWAS) approaches, at least 110 non-HLA genetic loci have been identified as associated with MS onset(25, 52-56). Most of these 110 variants' biological function still unclear, as yet, there are no benefits for MS treatment, based on GWAS result. Still, several are in regions of plausible relevance to MS, particularly those involved in immune function and those related to relevant environmental factors like vitamin D. Some of the most plausible will be summarised below.

Interleukin 7 receptor (IL7R)

IL7R was the only non-HLA MS common risk variants identified by the candidate gene method(57), and was subsequently confirmed by the first GWAS in multiple sclerosis(58). *IL7R* plays an important role in B and T-cell differentiation. Functionally, *IL7R* also involved in the regulation of innate and adaptive immunity, and the absence of this regulation can yield an inflammatory immune response that can damage the myelin sheath(57). *IL7* signalling is crucial for T-cell differentiation of $CD4^-$ $CD8^-$ thymocytes, $CD4^+$ $CD8^+$ cells survival, and immune homeostasis. The common variant, of the SNP rs6897932 in the *IL7R* gene, was allelic and functional associated with MS. This variant affects the brain and CNS, causing muscle weakness, poor coordination, numbness, and a variety of problems with the nervous system. The mode of effect for this SNP in neuropathology is potentially due to its role in differential transcription of the receptor protein. Beyond just impacting on disease risk, a recent study (270 cases, 303 controls) found that rs6897932 in *IL7R* was not only associated with MS susceptibility but also with disability progression in a Central European Slovak population: the allele C was a risk factor for MS, whereas the minor T allele was protective against both MS risk and against a more rapid disability progression (59).

Interleukin 2 receptor alpha *(IL2RA)*

There is a strong association between two variants of interleukin 2 receptor alpha (*IL2RA*) (OR: 1.19-1.25) and MS risk (58). *IL2RA* involved in regulation of T-cells and encoded the alpha chain of the interleukin-2 receptor. *IL2RA* chain also plays an

important role in MS mechanism because the *IL2RA* pathway involved in patients' adaptive and innate immune response(60). One study finds the functional explain for *IL2RA:* Granulocyte-macrophage colony-stimulating factor (GM-CSF) is strongly induced by interleukin 2 (*IL-2*). The polymorphisms in genes which in T helper (TH) cell differentiation pathway are associated with MS onset (61). Traboulsee and college found that *IL2RA* affected both MS onset and progression(62).

TNFRSF1A

TNFRSF1A encodes the p55 receptor for tumor necrosis factor alpha (*TNFα*). Previous studies show that dysregulation of the TNFα pathway may influence the risk of MS. The level of *TNFα* activity likely to be correlated with CNS inflammatory lesions. Studies also find that *TNFRSF1A* polymorphisms tended to have a second relapse within a year. *TNFRSF1A* not only associated with MS onset but also influenced the clinical course. One polymorphism (rs4149584) within *TNFRSF1A* cause an early MS onset and a slow MS progression(63).

CXCR5

Chemokine (C-X-C motif) receptor 5 also known as *CD185*. This gene encodes CXC chemokine receptor family proteins. *CXCR5* is one kind of protein coding gene. One case control study found the polymorphisms (rs630923) of *CXCR5* was associated with MS onset(64). In this study, rs630923 located into the transcription factor binding site of *CXCR5*(65).

TNFSF14

Tumor necrosis factor superfamily member 14 gene encodes the tumor necrosis factor(*TNF*) ligand family protein. This protein has been shown to stimulate the proliferation of T cells, which may influence the autoimmune response. One study with 477 MS cases and 481 control samples found the A allele of the polymorphism (rs1077667) in *TNFSF14* gene have a protective effect on MS onset, whereas the G allele was a risk allele for MS onset. In addition, *TNFSF14* was a microRNA target gene significantly decreased in MS. This result suggests that microRNA may influence the T-cell and molecules MS onset(66).

In addition, few relationships between these associated variants and MS progression have been identified. In our own studies, a prospective longitudinal cohort study of

persons who had a first clinical episode of demyelination (FDE), followed for 5-years. We found that seven non-HLA SNPs of the 110 variants predicted relapse and/or CDMS, and seven other non-HLA SNPs predicted the annualised change in disability status (ΔEDSS, as measured by Expanded Disability Status Scale). These results suggested that MS risk variants also influence MS clinical course (onset and progression).

4. Gene-environment interactions

Beyond genetic factors, many environmental factors have also been associated with MS onset and clinical course, such as vitamin D deficiency(67), smoking (68, 69) and EBV (Epstein-Barr virus) infection(70). Importantly, there is emerging evidence suggesting that MS may be caused by an interplay of multiple variants and environmental factors. Studies have indicated that gene-environment interactions can explain a part of the MS missing heritability: some environmental situation with particular genetic variants brings a significantly greater effect on MS onset and clinical course(6, 71, 72).

Vitamin D response elements(VDRE) located in the promoter region of *HLA-DRB1*15:01*, this locating suggests that different vitamin D situations might interact with HLA-DRB1 to influence MS onset(73). The VDRE modulates *HLA-DRB1* gene expression after stimulation with Vitamin D. In Australia, study with 466 MS cases and 498 controls showed that VDRE interact with different HLA-DRB1 promoter region polymorphism causes an 11 fold range risk of MS onset. However, other studies showed that other environmental factors interact with HLA risk variants play more role in MS onset. Similarly, the rate of progression of established MS is highly variable, e.g. monozygotic twins can have onsets of MS at significantly different ages, have completely different clinical presentations, and can progress at very different rates. This variability may be under genetic control or at least influenced by it. Currently, not many genes have been found that influence MS clinical course. The *HLA-DRB1*1501* locus decreases the age of onset. Therefore, it is likely that genes that interact with environmental factors may significantly influence the rate of MS clinical course and determine the different MS clinical course phenotype. It is thought that secondary progressive MS and PPMS represent the same pathological process and that MS relapses may represent a separate and additive inflammatory process more directly influenced by environmental determinants.

Lin and colleagues using Cox proportional hazards regression models to estimate the relationship between known MS common risk variants and MS relapse; and whether these SNPs influence the 25-hydroxyvitamin D (25(OH)D) –relapse association among 141 RRMS patients in Southern Tasmanian Multiple Sclerosis Longitudinal Study (the MSL study)(6). They identified five variants associated with MS relapse with significant cumulative genotype risk effects. They also found three SNPs which modified the relationship between the hazard of relapse and serum 25(OH)D levels. Unfortunately, no risk variants were identified for MS disability progression. This study indicated that gene and environment interactions may play an important role in MS clinical course. It is a mechanism by which MS clinical is driven. It will provide support for the role of serum 25(OH)D in MS onset and progression. Similarly, another study using the same cohort showed protein kinase C (PKC) family genes-25(OH)D interactions modulate MS clinical course(2). In this study, they identified two SNPs in *PRKCZ* and *PRKCH* interact with 25(OH)D levels to influence relapse. However, SNPs itself was not independently associated with hazard of relapse. They also found two SNPs within the *CYP2R1* and *PRKCB* were associated with 25(OH)D levels in relapse. *CYP2R1* has been previously identified to be associated with vitamin D levels in GWAS. The *PRKCB* gene has been previously associated with risk of rheumatoid arthritis and it is a member of the PKC gene family which is regulated by1,25 dihydroxyvitamin D (1,25(OH)D) in chondrocytes and mediated by *VDR* in downstream signalling pathways. These results suggest that gene-vitamin D interaction is associated with MS clinical course.

CYP24A1 and Vitamin D

rs6013897 ($p=6*10^{-10}$) in *CYP42A1* has been previously identified to be associated with vitamin D levels in GWAS (74). Vitamin D gene influence MS onset and clinical course by modulating vitamin D levels. Vitamin D deficiency has been considered as a risk factor for many complex diseases such as MS. Recent study found that the low vitamin D level plays an important role in MS risk. A large number of epidemiological data indicated that there is a high prevalence of MS in high-latitude and low ultraviolet radiation (UVR) cold area in contrast to the low prevalence of MS in low-latitude and high UVR warm area. This because Vitamin D3 is produced photochemically from 7-dehydrocholesterol in the skin by UVR(75). Then, vitamin D3 was absorbed by human body. Finally, vitamin D3 converted to 1,25-dihydroxyvitamin D ($1,25(OH)_2D_3$). Now, it was well established that Vitamin D in

humans mainly comes from sun exposure. Vitamin D play an essential role in the synthesis of myelin. Vitamin D as an environmental factor, not only influence whether a person will get MS but also impact the clinical course of MS. The involvement of the vitamin D in the onset and progression of MS was related to the essential role that vitamin D pathway plays in the autoimmune system. The immune system is regulated by combining of the active form vitamin D (Hydroxylates 25-hydroxyvitamin D) with the specific vitamin D receptor (*VDR*), and the function of *VDR* is regulated by its genetic structure. There are many restriction sites such as Bsm I and Apa I in VDR gene. One study suggests that the polymorphisms of the *VDR* gene may be associated with MS risk. Simpson and colleagues found that each 10nmol/l increase in vitamin D resulted in a nearly 12% reduction of MS relapse risk by using survival analysis(67). This is the first longitudinal cohort study to investigate the relationship between vitamin D and relapse of RRMS patients. The recent subsequent study shows evidence that Vitamin D plays an important role in myelin repair. This study finds vitamin D activated RXR gamma receptor protein and this protein was important in MS patients' myelin repairing (76). This finding functionally explained the relationship between the vitamin D level and progression of MS and indicated that vitamin D could be a target for the future myelin repair drug and a possible treatment for people with MS. In short, compared to other MS treatment, vitamin D based treatment are cheaper with fewer side effects. It is easier for people to increase vitamin D level by sun exposure and daily intake.

Many research results bring the question that why a number of genetically susceptible individuals maintain healthy while others develop into the disease. One potential answer is genetic heterogeneity, and the other is gene-environmental interactions. Genetic epidemiology implies that genetic background has an important complementary role in MS onset. If genetic factors keep the same in individuals, the environment determines the threshold of MS onset(77).

5. Epigenetic

Recently, major progress has suggested that epigenetic mechanism contributes to the pathophysiology of MS. These partly mediate the response to environmental influences and changes in gene expression. Epigenetics is known as heritable changes in gene expression yet not modifying underlying DNA sequence(78).In multiple

sclerosis, epigenetic changes mediate the response to environmental factors (5, 79) and then change gene's expression. The epigenetic mechanisms consist of histone modification, DNA methylation and miRNA-associated post-transcriptional gene silencing(80). The import finding of epigenetic processes in cells are DNA methylation and histone deacetylation. MS usually transmitted to children more by female than male implies an epigenetic contribution(81). The potential reason for this parent effect might be the *HLA-DRB1*15:01* allele, which is the major genetic risk factor for MS and regulated by epigenetic such as DNA methylation, histone deacetylation, and miRNA-associated silencing.

Epigenetic changes can explain much of heterogeneity in MS clinical course. For example, gene methylation of MS patient's blood study found DNA methylation may be used as markers of MS disease activity (82). Study on methylation changes in cancer patients 56 genes show that 15 genes in the cell-free plasma DNA occur methylation in MS patients and healthy controls group, further, that 5 in 15 genes promote methylation of genes may also distinguishing patients with remission and exacerbation. Another study showed thatmiRNA-572 was significantly upregulated and downregulated during MS clinical course such as disease relapse and remission phases. This result indicated that miRNA-572 can be a biomarker for remyelination(83). These results suggest that all these microRNAs might be potential targets for the treatment of MS.

The epigenetics contribution to MS is just beginning to come to light, several mechanisms of epigenetic change in patients with MS were identified. Believing that there will be more findings in the near future. It has opened a new era to us in MS research. DNA demethylation of epigenetic, histone deacetylation and regulation mechanism of miRNA opened up enticing prospect of new therapies for MS. There is a vast potential to use epigenetics for both MS preventive treatments and personalised therapies. It also brings a great hope to develop complex diseases' molecular targeted genome research(84, 85). Epigenetics became the newest MS genetic research method in future.

6. Challenges for future research in MS genetic epidemiology

Clinical application: Translational medicine- translating predictive factors (GWAS results, epigenetic results) to clinical application. Let doctor and patient to know genetic information easily.

Biological implications: To test whether GWAS results indicated new mechanisms of MS. Identify potential polymorphism targets, gene coding functions, and epigenetic regulation.

Large magnitude: The effect of one non-HLA common variant were small. (OR ranging from 1.1-1.5). There has been a cumulative effect of many variants into a cumulative genetic risk score for greater clinical utility.

7. Conclusions and future directions

During last decades, major progress has been made in identifying MS susceptibility genes. In particular, the role of HLA loci has been fairly well substantiated, but recently other non-HLA genes, including those involved in immune function and vitamin D metabolism, among others, have been indicated. However, while the role of these loci in MS onset has made great strides, there has been comparatively less success in examining these factors' role in MS clinical course. While there are methodological hindrances to success in this area, it may also be due to the complexity and inter-patient heterogeneity in disease progression, to say nothing of the potential confounding impacts of post-onset behavioural changes and disease-modifying therapies. All that said, there is hope for the future. The new discovery will depend on large biological samples, comprehensive databases of environmental/behavioural, genetic and clinical factors, new genetic technologies and statistical models. Further research of MS susceptibility genes may lead to a better understanding of MS susceptibility & pathogenesis, and potentially to a more personalised medical treatment approach.

References

1. Lassmann H, Bruck W, Lucchinetti CF. The immunopathology of multiple sclerosis: an overview. Brain Pathol. 2007;17(2):210-8.
2. Lin R, Taylor BV, Simpson S, Jr., Charlesworth J, Ponsonby AL, Pittas F, et al. Novel modulating effects of PKC family genes on the relationship between serum vitamin D and relapse in multiple sclerosis. J Neurol Neurosurg Psychiatry. 2014;85(4):399-404.

3. Compston A, Coles A. Multiple sclerosis. Lancet. 2008;372(9648):1502-17.
4. Pelletier D, Hafler DA. Fingolimod for multiple sclerosis. N Engl J Med. 2012;366(4):339-47.
5. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part I: the role of infection. Ann Neurol. 2007;61(4):288-99.
6. Lin R, Taylor BV, Simpson S, Jr., Charlesworth J, Ponsonby AL, Pittas F, et al. Association between multiple sclerosis risk-associated SNPs and relapse and disability--a prospective cohort study. Mult Scler. 2014;20(3):313-21.
7. Sawcer S, Franklin RJ, Ban M. Multiple sclerosis genetics. Lancet Neurol. 2014;13(7):700-9.
8. Psychiatric GCBDWG. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat Genet. 2011;43(10):977-83.
9. Pulst SM. Genetic linkage analysis. Arch Neurol. 1999;56(6):667-72.
10. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nature reviews Genetics. 2005;6(2):95-108.
11. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic acids research. 2014;42(Database issue):D1001-6.
12. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. Nature reviews Genetics. 2010;11(7):459-63.
13. Li MX, Yeung JM, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. Hum Genet. 2012;131(5):747-56.
14. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300.
15. Manolio TA, Bailey-Wilson JE, Collins FS. Genes, environment and the value of prospective cohort studies. Nature reviews Genetics. 2006;7(10):812-20.
16. Sackett DL. Bias in analytic research. J Chronic Dis. 1979;32(1-2):51-63.
17. Ponsonby AL, Lucas RM, Dear K, van der Mei I, Taylor B, Chapman C, et al. The physical anthropometry, lifestyle habits and blood pressure of people presenting with a first clinical demyelinating event compared to controls: The Ausimmune study. Mult Scler. 2013;19(13):1717-25.
18. Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, et al. Gene map of the extended human MHC. Nat Rev Genet. 2004;5(12):889-99.
19. Bertrams HJ, Kuwert EK. Association of histocompatibility haplotype HLA-A3-B7 with multiple sclerosis. J Immunol. 1976;117(5 Pt.2):1906-12.
20. Wadia NH, Trikannad VS, Krishnaswamy PR. HLA antigens in multiple sclerosis amongst Indians. J Neurol Neurosurg Psychiatry. 1981;44(9):849-51.
21. Hollenbach JA, Oksenberg JR. The immunogenetics of multiple sclerosis: A comprehensive review. J Autoimmun. 2015;64:13-25.
22. Olerup O, Hillert J. HLA class II-associated genetic susceptibility in multiple sclerosis: a critical evaluation. Tissue Antigens. 1991;38(1):1-15.
23. Oksenberg JR, Barcellos LF, Cree BA, Baranzini SE, Bugawan TL, Khan O, et al. Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. Am J Hum Genet. 2004;74(1):160-7.

24. McElroy JP, Cree BA, Caillier SJ, Gregersen PK, Herbert J, Khan OA, et al. Refining the association of MHC with multiple sclerosis in African Americans. Hum Mol Genet. 2010;19(15):3080-8.
25. International Multiple Sclerosis Genetics C, Wellcome Trust Case Control C, Sawcer S, Hellenthal G, Pirinen M, Spencer CC, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011;476(7359):214-9.
26. Stankovich J, Butzkueven H, Marriott M, Chapman C, Tubridy N, Tait BD, et al. HLA-DRB1 associations with disease susceptibility and clinical course in Australians with multiple sclerosis. Tissue Antigens. 2009;74(1):17-21.
27. Dyment DA, Ebers GC, Sadovnick AD. Genetics of multiple sclerosis. Lancet Neurol. 2004;3(2):104-10.
28. Modin H, Olsson W, Hillert J, Masterman T. Modes of action of HLA-DR susceptibility specificities in multiple sclerosis. Am J Hum Genet. 2004;74(6):1321-2.
29. Fukazawa T, Yamasaki K, Ito H, Kikuchi S, Minohara M, Horiuchi I, et al. Both the HLA-CPB1 and -DRB1 alleles correlate with risk for multiple sclerosis in Japanese: clinical phenotypes and gender as important factors. Tissue Antigens. 2000;55(3):199-205.
30. Yoshimura S, Isobe N, Yonekawa T, Matsushita T, Masaki K, Sato S, et al. Genetic and infectious profiles of Japanese multiple sclerosis patients. PLoS One. 2012;7(11):e48592.
31. Field J, Browning SR, Johnson LJ, Danoy P, Varney MD, Tait BD, et al. A polymorphism in the HLA-DPB1 gene is associated with susceptibility to multiple sclerosis. PLoS One. 2010;5(10):e13454.
32. Wisniewski A, Bilinska M, Klimczak A, Wagner M, Majorczyk E, Nowak I, et al. Association of the HLA-G gene polymorphism with multiple sclerosis in a Polish population. Int J Immunogenet. 2010;37(4):307-11.
33. Alcina A, Abad-Grau Mdel M, Fedetz M, Izquierdo G, Lucas M, Fernandez O, et al. Multiple sclerosis risk variant HLA-DRB1*1501 associates with high expression of DRB1 gene in different human populations. PLoS One. 2012;7(1):e29819.
34. Wu JS, Qiu W, Castley A, James I, Joseph J, Christiansen FT, et al. Presence of CSF oligoclonal bands (OCB) is associated with the HLA-DRB1 genotype in a West Australian multiple sclerosis cohort. J Neurol Sci. 2010;288(1-2):63-7.
35. Qiu W, Wu JS, Castley A, James I, Joseph J, Christiansen FT, et al. Clinical profile and HLA-DRB1 genotype of late onset multiple sclerosis in Western Australia. J Clin Neurosci. 2010;17(8):1009-13.
36. Ramagopalan SV, Byrnes JK, Dyment DA, Guimond C, Handunnetthi L, Disanto G, et al. Parent-of-origin of HLA-DRB1*1501 and age of onset of multiple sclerosis. J Hum Genet. 2009;54(9):547-9.
37. Chao MJ, Herrera BM, Ramagopalan SV, Deluca G, Handunetthi L, Orton SM, et al. Parent-of-origin effects at the major histocompatibility complex in multiple sclerosis. Hum Mol Genet. 2010;19(18):3679-89.
38. Ebers GC, Bulman DE, Sadovnick AD, Paty DW, Warren S, Hader W, et al. A population-based study of multiple sclerosis in twins. N Engl J Med. 1986;315(26):1638-42.
39. Kuusisto H, Kaprio J, Kinnunen E, Luukkaala T, Koskenvuo M, Elovaara I. Concordance and heritability of multiple sclerosis in Finland: study on a nationwide series of twins. Eur J Neurol. 2008;15(10):1106-10.

40. Ristori G, Cannoni S, Stazi MA, Vanacore N, Cotichini R, Alfo M, et al. Multiple sclerosis in twins from continental Italy and Sardinia: a nationwide study. Ann Neurol. 2006;59(1):27-34.
41. Hansen T, Skytthe A, Stenager E, Petersen HC, Bronnum-Hansen H, Kyvik KO. Concordance for multiple sclerosis in Danish twins: an update of a nationwide study. Mult Scler. 2005;11(5):504-10.
42. Wu JS, James I, Qiu W, Castley A, Christiansen FT, Carroll WM, et al. HLA-DRB1 allele heterogeneity influences multiple sclerosis severity as well as risk in Western Australia. J Neuroimmunol. 2010;219(1-2):109-13.
43. Silva AM, Pereira C, Bettencourt A, Carvalho C, Couto AR, Leite MI, et al. The role of HLA-DRB1 alleles on susceptibility and outcome of a Portuguese Multiple Sclerosis population. J Neurol Sci. 2007;258(1-2):69-74.
44. Weatherby SJ, Thomson W, Pepper L, Donn R, Worthington J, Mann CL, et al. HLA-DRB1 and disease outcome in multiple sclerosis. J Neurol. 2001;248(4):304-10.
45. Romero-Pinel L, Pujal JM, Martinez-Yelamos S, Gubieras L, Matas E, Bau L, et al. Epistasis between HLA-DRB1 parental alleles in a Spanish cohort with multiple sclerosis. J Neurol Sci. 2010;298(1-2):96-100.
46. Barcellos LF, Oksenberg JR, Begovich AB, Martin ER, Schmidt S, Vittinghoff E, et al. HLA-DR2 dose effect on susceptibility to multiple sclerosis and influence on disease course. Am J Hum Genet. 2003;72(3):710-6.
47. Barcellos LF, Sawcer S, Ramsay PP, Baranzini SE, Thomson G, Briggs F, et al. Heterogeneity at the HLA-DRB1 locus and risk for multiple sclerosis. Hum Mol Genet. 2006;15(18):2813-24.
48. Perini P, Tagliaferri C, Belloni M, Biasi G, Gallo P. The HLA-DR13 haplotype is associated with "benign" multiple sclerosis in northeast Italy. Neurology. 2001;57(1):158-9.
49. Wu JS, James I, Wei Q, Castley A, Christiansen FT, Carroll WM, et al. Influence of HLA-DRB1 allele heterogeneity on disease risk and clinical course in a West Australian MS cohort: a high-resolution genotyping study. Mult Scler. 2010;16(5):526-32.
50. Okuda DT, Srinivasan R, Oksenberg JR, Goodin DS, Baranzini SE, Beheshtian A, et al. Genotype-Phenotype correlations in multiple sclerosis: HLA genes influence disease severity inferred by 1HMR spectroscopy and MRI measures. Brain. 2009;132(Pt 1):250-9.
51. Cree BA, Reich DE, Khan O, De Jager PL, Nakashima I, Takahashi T, et al. Modification of Multiple Sclerosis Phenotypes by African Ancestry at HLA. Arch Neurol. 2009;66(2):226-33.
52. International Multiple Sclerosis Genetics C, Beecham AH, Patsopoulos NA, Xifara DK, Davis MF, Kemppinen A, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat Genet. 2013;45(11):1353-60.
53. International Multiple Sclerosis Genetics C, International IBDGC, International IBDGCI. Class II HLA interactions modulate genetic risk for multiple sclerosis. Nat Genet. 2015.
54. Baranzini SE, Wang J, Gibson RA, Galwey N, Naegelin Y, Barkhof F, et al. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. Hum Mol Genet. 2009;18(4):767-78.
55. Martinelli-Boneschi F, Esposito F, Brambilla P, Lindstrom E, Lavorgna G, Stankovich J, et al. A genome-wide association study in progressive multiple sclerosis. Mult Scler. 2012;18(10):1384-94.

56. Patsopoulos NA, Barcellos LF, Hintzen RQ, Schaefer C, van Duijn CM, Noble JA, et al. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. PLoS Genet. 2013;9(11):e1003926.
57. Gregory SG, Schmidt S, Seth P, Oksenberg JR, Hart J, Prokop A, et al. Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis. Nat Genet. 2007;39(9):1083-91.
58. International Multiple Sclerosis Genetics C, Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, et al. Risk alleles for multiple sclerosis identified by a genomewide study. N Engl J Med. 2007;357(9):851-62.
59. Cierny D, Hanysova S, Michalik J, Kantorova E, Kurca E, Skerenova M, et al. Genetic variants in interleukin 7 receptor alpha chain (IL-7Ra) are associated with multiple sclerosis risk and disability progression in Central European Slovak population. J Neuroimmunol. 2015;282:80-4.
60. Hoppenbrouwers IA, Hintzen RQ. Genetics of multiple sclerosis. Biochim Biophys Acta. 2011;1812(2):194-201.
61. Hartmann FJ, Khademi M, Aram J, Ammann S, Kockum I, Constantinescu C, et al. Multiple sclerosis-associated IL2RA polymorphism controls GM-CSF production in human TH cells. Nat Commun. 2014;5:5056.
62. Traboulsee AL, Bernales CQ, Ross JP, Lee JD, Sadovnick AD, Vilarino-Guell C. Genetic variants in IL2RA and IL7R affect multiple sclerosis disease risk and progression. Neurogenetics. 2014;15(3):165-9.
63. Comabella M, Caminero AB, Malhotra S, Agullo L, Fernandez O, Reverter F, et al. TNFRSF1A polymorphisms rs1800693 and rs4149584 in patients with multiple sclerosis. Neurology. 2013;80(22):2010-6.
64. International Multiple Sclerosis Genetics C, Lill CM, Schjeide BM, Graetz C, Ban M, Alcina A, et al. MANBA, CXCR5, SOX8, RPS6KB1 and ZBTB46 are genetic risk loci for multiple sclerosis. Brain. 2013;136(Pt 6):1778-82.
65. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.
66. Jernas M, Malmestrom C, Axelsson M, Nookaew I, Wadenvik H, Lycke J, et al. MicroRNA regulate immune pathways in T-cells in multiple sclerosis (MS). BMC Immunol. 2013;14:32.
67. Simpson S, Jr., Taylor B, Blizzard L, Ponsonby AL, Pittas F, Tremlett H, et al. Higher 25-hydroxyvitamin D is associated with lower relapse risk in multiple sclerosis. Ann Neurol. 2010;68(2):193-203.
68. Roudbari SA, Ansar MM, Yousefzad A. Smoking as a risk factor for development of Secondary Progressive Multiple Sclerosis: A study in IRAN, Guilan. J Neurol Sci. 2013;330(1-2):52-5.
69. Pittas F, Ponsonby AL, van der Mei IA, Taylor BV, Blizzard L, Groom P, et al. Smoking is associated with progressive disease course and increased progression in clinical disability in a prospective cohort of people with multiple sclerosis. J Neurol. 2009;256(4):577-85.
70. Lincoln JA, Cook SD. An overview of gene-epigenetic-environmental contributions to MS causation. J Neurol Sci. 2009;286(1-2):54-7.
71. Zhou Y, Zhu G, Charlesworth JC, Simpson S, Jr., Rubicz R, Goring HH, et al. Genetic loci for Epstein-Barr virus nuclear antigen-1 are associated with risk of multiple sclerosis. Mult Scler. 2016.
72. Ramagopalan SV, Maugeri NJ, Handunnetthi L, Lincoln MR, Orton SM, Dyment DA, et al. Expression of the multiple sclerosis-associated MHC class II Allele HLA-DRB1*1501 is regulated by vitamin D. PLoS Genet. 2009;5(2):e1000369.

73. Cocco E, Meloni A, Murru MR, Corongiu D, Tranquilli S, Fadda E, et al. Vitamin D responsive elements within the HLA-DRB1 promoter region in Sardinian multiple sclerosis associated alleles. PLoS One. 2012;7(7):e41678.
74. Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, Berry D, et al. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. Lancet. 2010;376(9736):180-8.
75. Holick MF. Vitamin D: A millenium perspective. J Cell Biochem. 2003;88(2):296-307.
76. de la Fuente AG, Errea O, van Wijngaarden P, Gonzalez GA, Kerninon C, Jarjour AA, et al. Vitamin D receptor-retinoid X receptor heterodimer signaling regulates oligodendrocyte progenitor cell differentiation. J Cell Biol. 2015;211(5):975-85.
77. Ebers GC. Environmental factors and multiple sclerosis. Lancet Neurol. 2008;7(3):268-77.
78. Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. Nature. 2004;429(6990):457-63.
79. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part II: Noninfectious factors. Ann Neurol. 2007;61(6):504-13.
80. Koch MW, Metz LM, Kovalchuk O. Epigenetic changes in patients with multiple sclerosis. Nat Rev Neurol. 2013;9(1):35-43.
81. Ebers GC, Sadovnick AD, Dyment DA, Yee IM, Willer CJ, Risch N. Parent-of-origin effect in multiple sclerosis: observations in half-siblings. Lancet. 2004;363(9423):1773-4.
82. Liggett T, Melnikov A, Tilwalli S, Yi Q, Chen H, Replogle C, et al. Methylation patterns of cell-free plasma DNA in relapsing-remitting multiple sclerosis. J Neurol Sci. 2010;290(1-2):16-21.
83. Mancuso R, Hernis A, Agostini S, Rovaris M, Caputo D, Clerici M. MicroRNA-572 expression in multiple sclerosis patients with different patterns of clinical progression. J Transl Med. 2015;13:148.
84. Zhou Y, Simpson S, Jr., Holloway AF, Charlesworth J, van der Mei I, Taylor BV. The potential role of epigenetic modifications in the heritability of multiple sclerosis. Mult Scler. 2014;20(2):135-40.
85. Kucukali CI, Kurtuncu M, Coban A, Cebi M, Tuzun E. Epigenetics of multiple sclerosis: an updated review. Neuromolecular Med. 2015;17(2):83-96.